This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

(19) World Intellectual Property Organization International Bureau



| 1881 | 1880 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 | 1881 |

(43) International Publication Date 31 May 2001 (31.05.2001)

PCT

(10) International Publication Number WO 01/38360 A2

(51) International Patent Classification⁷: C07K 14/005

(21) International Application Number: PCT/US00/32326

(22) International Filing Date:

22 November 2000 (22.11.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/167,502

24 November 1999 (24.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:

US Filed on 60/167,502 (CIP) 24 November 1999 (24.11.1999)

(71) Applicant (for all designated States except US): CHI-RON CORPORATION [US/US]; 4560 Horton Street,

(72) Inventors; and

Emeryville, CA 94608 (US).

(75) Inventors/Applicants (for US only): COIT, Doris [US/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US). MEDINA-SELBY, Angelica [CL/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US). SELBY, Mark [US/US]; Chiron Corporation, 4560 Horton Street, R-440,

Emeryville, CA 94608 (US). **HOUGHTON, Michael** [GB/US]; Chiron Corporation, 4560 Horton Street, R-440, Emeryville, CA 94608 (US).

- (74) Agents: HARBIN, Alisa, A.; Chiron Corporation, Intellectual Property R440, P.O. Box 8097, Emeryville, CA 94662-8097 et al. (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

 Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HCV NON-STRUCTURAL POLYPEPTIDE

(57) Abstract: Polypeptides comprising a mutant non-structural Hepatitis C virus useful in diagnostic and/or immunogenic compositions are disclosed, in which the mutant is an N-terminal mutation that functionally disrupt the catalytic domain of NS3. Polynucleotides encoding these polypeptides, host cells transformed with polynucleotides and methods of using the polypeptides and polynucleotides are also disclosed.

NOVEL HCV NON-STRUCTURAL POLYPEPTIDE

FIELD OF THE INVENTION

The present invention relates to polypeptides comprising a mutant nonstructural Hepatitis C virus ("HCV") polypeptide useful for immunogenic compounds for use against HCV, methods of preparing and using the same, and immunogenic compositions comprising the same. The present invention also relates to compositions comprising (a) a mutant non-structural HCV polypeptide and (b) a viral polypeptide that is not a non-structural HCV polypeptide and methods of using these compositions.

15

5

10

BACKGROUND OF THE INVENTION

HCV is now recognized as the major agent of chronic hepatitis and liver disease worldwide. It is estimated that HCV infects about 400 million people worldwide, corresponding to more than 3% of the world population.

20

25

Hepatitis C virus ("HCV") is a small enveloped RNA *flavivirus*, which contains a positive-stranded RNA genome of about 10 kilobases. The genome has a single uninterrupted ORF that encodes a protein of 3010-3011 amino acids. The structural proteins of HCV include a core protein (C), which is highly immunogenic, as well as two envelope proteins (E1 and E2), which likely form a heterodimer *in vivo*, and non-structural proteins NS2-NS5. It is known that the NS3 region of the virus is important for post-translational processing of the polyprotein into individual proteins, and the NS5 region encodes an RNA-dependant RNA polymerase.

30

Virus-specific T lymphocytes, along with neutralizing antibodies, are the mainstay of the antiviral immune defense in established viral infections. Whereas CD8⁺ cytotoxic T cells eliminate virus-infected-cells, CD4⁺ T helper cells are essential for the efficient regulation of the antiviral immune response. CD4⁺ T helper cells recognize specific antigens as peptides bound to autologous HLA class II molecules (viral antigens or particles are taken up by professional antigen-presenting cells, processed to peptides, bound to HLA class II molecules in the lysosomal compartment,

and transported back to the cell surface). Several observations support an important role of CD4⁺ T cells in the elimination of HCV infection. Tsai *et al.*, 1997 Hepatology 25:449-458; Diepolder et al 1995 Lancet 346: 1—6-1009; Missale et al 1996 JCI 98: 706-714; Botarelli et al 1993; Gastro 104: 580-587; Diepolder et al 1997 J.Virol 71: 6011. Immunogenic peptides usually have a minimal length of 8-11 amino acids. However, since the peptide binding groove of HLA class II molecules seems to be open at both ends, longer peptides are tolerated. Thus peptides eluted from HLA class II molecules are extremely polymorphic and each allele seems to have its individual requirements for peptide binding. Thus the HLA class II repertoire of a given individual determines which viral peptides can be presented to T cells. Recognition of the specific HLA-peptide complex by the T cell receptor accompanied by appropriate costimulatory signals lead to T cell activation, secretion of cytokines, and T cell proliferation.

5

10

15

20

25

30

Numerous studies demonstrate that HLA Class II restricted CD4⁺ responses are determined by stimulating peripheral blood mononuclear cells with recombinant viral antigens or peptides. Botarelli *et al.*, (1993) Gastroenterology 104:580-587; Farrari *et al.*, (1994) Hepatology 19:286-295; Minutello *et al.*, (1993) C. J. Exp. Med. 178:17-25; Hoffmann *et al.*, (1995) Hepatology 21:632-638; Iwata *et al.*, (1995) Hepatology 22:1057-1064; and Tsai.*et al.*, (1995) Hepatology 21:908-912.

Polyclonal multispecific CD8⁺ T cell responses have been detected in patients with chronic hepatitis C. Additionally, CD8⁺ CTL's were shown to be important in resolving acute HCV infection in chimpanzees (Cooper *et al.*, Immunity 1999). About 50% of patients with chronic hepatitis C demonstrate a detectable virus-specific CD4⁺ T cell response, which is most frequently directed against HCV core and/or NS4 and tends to be more common in patients who achieve sustained viral clearance during interferon-α therapy.

Depending on the pattern of lymphokines, CD4⁺ T helper cells have been classified as TH1, TH0, or TH2. Cytokines of the TH1 type are typically IFN-γ, lymphotoxin, and interleukin-2 (IL-2), which are believed to support activation of virus-specific CD8⁺ T cells and natural killer cells. The TH2 cytokines IL-4, IL-5, IL-10, and IL-13 are important for B cell activation and differentiation, thus inducing a humoral immune response.

During acute hepatitis C infection a strong and sustained TH1/TH0 response to NS3 and possibly to other nonstructural proteins is associated with a self-limited course of the disease. Diapolder *et al.*, (1995) Lancet 346:1006-1007, showed all CD4⁺ T cell clones to have a TH1 or TH0 cytokine profile, suggesting that the clones support cytotoxic immune mechanisms *in vivo*. The majority of CD4⁺ T cell clones responded to a relatively short segment of NS3, namely amino acids 1207-1278, suggesting that this region of NS3 is immunodominant for CD4⁺ T cells. More than 70% of those who contract HCV develop chronic infection and hepatitis, and a significant portion of them progress to cirrhosis and eventually hepatocellular carcinoma. The only approved therapy at present is a 6- to 12- month course of interferon α, which leads to sustained improvement in only 20% of patients. So far, no commercial vaccine is available.

Thus, there remains a need for compositions and methods capable of promoting anti-HCV responses.

15 SUMMARY OF THE INVENTION

5

10

20

25

30

In one aspect, the present invention relates to isolated polypeptides comprising mutant hepatitis C ("HCV") polypeptides comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, NS3 is encoded by a nucleic acid sequence having an N-terminal deletion to remove the catalytic domain. The NS mutant polypeptides can include NS3, NS4s, NS4b, NS5a, NS5b or portions thereof. For example, in various embodiments, the mutant NS polypeptide comprises NS3, NS4 (NS4a and NS4b) and NS5 (NS5a and NS5b). In other embodiments, the NS polypeptide consists of NS3 and NS4 (for example, NS4a and/or NS4b) or NS3 and NS5 (for example, NS5a and/or NS5b). Other combinations of full-length or fragments of non-structural components are also contemplated.

In another preferred aspect, the polypeptides further comprise a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV.

Thus, the invention includes an isolated mutant non-structural ("NS") HCV polypeptide comprising a polypeptide having a mutation in the catalytic domain of NS3 that functionally disrupts the catalytic domain. The mutation can be, for example, a deletion or a substitution mutation. In certain embodiments, the mutant NS polypeptide comprises NS3, NS4 and NS5. In other embodiments, the mutant NS polypeptides described herein further comprise a second viral polypeptide that is not NS3, NS4, or NS5 of HCV, for example an HCV Core polypeptide ("C"), or fragment thereof, or an HCV envelope protein ("E"), for example E1 and/or E2. In certain embodiments, C is truncated (e.g., at amino acid 121).

5

10

15

20

25

30

In another aspect, the present invention relates to compositions comprising any of the mutant hepatitis C ("HCV") polypeptides described herein, for example polypeptides comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, NS3 is encoded by a nucleic acid sequence having an N-terminal deletion to disrupt the function of the catalytic domain, for example by removing this domain. In another preferred aspect, the polypeptides further comprise a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another aspect, the invention includes a composition comprising (a) any of the polypeptides described herein; and (b) a pharmaceutically acceptable excipient (e.g., carrier and/or adjuvant).

In another aspect, the invention includes an isolated and purified polynucleotide which encodes any of the mutant HCV polypeptides described herein. In certain embodiments, the invention includes a composition comprising (a) the isolated purified polynucleotide encoding any of the mutant HCV polypeptides; and (b) a pharmaceutically acceptable excipient. The polynucleotide, can be for example, DNA in a plasmid, or is in a plasmid. Additionally, the polynucleotides described herein may be included in an expression vector as shown in the attached Figures and Sequence Listings.

In another aspect, the present invention relates to host cells transformed with expression vectors comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising at least portions of NS3, NS4, and NS5. In a preferred aspect, the expression vectors of the host cells further comprises at least one nucleic acid sequence encoding a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another preferred aspect the nucleic acid sequences of the expression vectors are coexpressed. In yet another preferred aspect, the host cells are yeast cells or mammalian cells.

5

10

15

20

25

30

In another aspect, the present invention relates to expression vectors comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising NS3, NS4, and NS5. In a preferred aspect, the expression vectors of the host cells further comprises at least one nucleic acid sequence encoding a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Importantly, such polypeptides need not be encoded by a natural HCV genome, such as, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another aspect, the present invention relates to methods of preparing a mutant HCV polypeptides. In a preferred aspect, the method comprises the steps of transforming a host cell with an expression vector, said vector comprising a nucleic acid sequence encoding a mutant HCV polypeptide comprising at least portions of NS3, NS4, and NS5, and isolating said polypeptide. In another preferred aspect the HCV polypeptide further comprises a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably. truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by

a natural HCV genome, and include, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, such as, for example, polypeptides of HBV. In another preferred aspect the host cells are yeast cells or mammalian cells.

In another aspect, the present invention relates to antibodies which specifically bind to mutant HCV polypeptide comprising NS3, NS4, and NS5, and to methods of making and using the same. In a preferred aspect, the HCV polypeptide further comprises a viral polypeptide that is not a non-structural HCV polypeptide. Such polypeptides are preferably C, or antigenic fragments thereof, more preferably, truncated C of HCV. Other polypeptides are preferably E, or antigenic fragments thereof, more preferably, E1 or E2 of HCV. Such polypeptides need not be encoded by a natural HCV genome, such as, for example, truncated or otherwise mutant HCV polypeptides or polypeptides derived from other genomes, and include, for example, polypeptides of HBV. In another preferred aspect, the antibody is either monoclonal or polyclonal.

In yet another aspect, a method of preparing a mutant NS HCV polypeptide, wherein the method comprises the steps of (a) transforming a host cell with any of the expression vectors described herein, under conditions wherein the polypeptide is expressed; and (b) isolating the polypeptide. The host cell can be, for example, a yeast cell, a mammalian cell a plant cell or an insect cell. The polypeptide can be expressed and isolated intracellularly or can be secreted and isolated from the surrounding environment.

In a still further aspect, a method of eliciting an immune response in a subject is provided. The immune response can be elicited by administering any of the polynucleotides and/or polypeptides described herein in one or multiple doses.

These and other embodiments of the subject invention will readily occur to those of skill in the art in light of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

5

10

15

20

25

- FIG. 1 shows the cloning scheme for generating pCMV-NS35.
 - FIG. 2 shows the 9621bp vector pCMV-NS35.

FIG. 3 shows the nucleic acid sequence of pCMV-NS35 (SEQ ID NO:1), including the nucleic acid sequence of the NS35 ORF, and also the translation of NS35 (SEQ ID NO:2).

- FIG. 4 shows the 9621bp pCMV-delNS35.
- 5 FIG. 5 shows the nucleic acid sequence of pCMV-delNS35 (SEQ ID NO:3), including the nucleic acid sequence of the delNS35 ORF, and also the translation of the delNS35 polypeptide (SEQ ID NO:4).
 - FIG. 6 shows the 4276bp pCMV-II.
 - FIG. 7 shows the nucleic acid sequence of pCMV-II (SEQ ID NO:5).
- 10 FIG. 8 shows the 6300bp pCMV-NS34A.

30

- FIG. 9 shows the nucleic acid sequence of pCMV-NS34A (SEQ ID NO:6), including the nucleic acid sequence of the NS34A ORF, and also the translation of NS34A (SEQ ID NO:7).
- FIG. 10 shows the cloning scheme for generating pd.ΔNS3NS5.
- 15 FIG. 11 shows the nucleic and amino acid sequences of pd.ΔNS3NS5 (SEQ ID NO:8 and 9).
 - FIG. 12 shows the Western blot of proteins expressed by S. cerevisiae strain AD3 transformed with pd.ΔNS3NS5.
 - FIG. 13 shows the cloning scheme for generating pd. ΔNS3NS5.pj.
- FIG. 14 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj (SEQ ID NO:10 and 11).
 - FIG. 15 shows the Western blot of proteins expressed by S. cerevisiae strain AD3 transformed with pd. Δ NS3NS5.pj, specifically demonstrating the expression of Δ NS3NS5 polypeptide.
- FIG. 16 shows the cloning scheme for generating pdΔNS3NS5.pj.core121RT and pdΔNS3NS5.pj.core173RT.
 - FIG. 17 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core121 (SEQ ID NO:12 and 13).
 - FIG. 18 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core173 (SEQ ID NO:14 and 15).
 - FIG. 19 shows the Western blot of proteins expressed by S. cerevisiae strain AD3 transformed with pd. ΔNS3NS5.pj, specifically demonstrating the expression of

ΔNS3NS5.core121 and ΔNS3NS5.core173 polypeptides. Lanes 1 and 7 show See Blue Standards. Lane 2 shows control yeast plasmid. Lanes 3 and 4 show ΔNS3NS5.core121RT polypeptide, colonies 1 and 2. Lanes 5 and 6 show ΔNS3NS5.core173RT polypeptide, colonies 3 and 4.

- 5 FIG. 20 shows the cloning scheme for generating pdΔNS3NS5.pj.core140RT and pdΔNS3NS5.pj.core150RT.
 - FIG. 21 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core140 (SEQ ID NO:16 and 17).
- FIG. 22 shows the nucleic and amino acid sequences of pd.ΔNS3NS5.pj.core150 (SEQ 10 NO:18 and 19).
 - FIG. 23 shows the Western blot of proteins expressed by S. cerevisiae strain AD3 transformed with pd.ΔNS3NS5.pj, specifically demonstrating the expression of ΔNS3NS5core140 and ΔNS3NS5core150 polypeptides. Lane 1 shows See Blue Standards. Lanes 2 and 3 show ΔNS3NS55core140RT polypeptide, colonies 5 and 6.
- Lanes 4 and 5 show ΔNS3NS5core150RT polypeptide, colonies 7 and 8. Lane 6 shows control yeast plasmid. Lane 7 shows ΔNS3NS5core121RT polypeptide, colony 1. Lane 8 shows ΔNS3NS5core173RT polypeptide, colony 5.

DETAILED DESCRIPTION OF THE INVENTION

- The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA techniques, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, et al., MOLECULAR CLONING; A LABORATORY MANUAL (1989); DNA CLONING, VOLUMES I AND II (D. N.
- Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M. J. Gait ed., 1984);
 NUCLEIC ACID HYBRIDIZATION (B. D. Hames & S. J. Higgins eds. 1984);
 TRANSCRIPTION AND TRANSLATION (B. D. Hames & S. J. Higgins eds. 1984);
 ANIMAL CELL CULTURE (R. I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR
 CLONING (1984); the series, METHODS OF ENZYMOLOGY (Academic Press,
- Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J. H. Miller and M. P. Calos eds. 1987, Cold Springs Harbor Laboratory), Methods in Enzymology Vol.

154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively); Mayer and Walker eds. (1987), IMMUNOHISTOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London); Scopes, (1987), PROTEIN PURIFICATION: PRINCIPALS AND PRACTICE, Second Edition (Springer-Verlag, New York); and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D. M. Weir and C. C. Blackwell eds. 1986).

It must be noted that, as used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more antigens, and the like.

I. Definitions

5

10

15

20

25

30

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

The term "hepatitis C virus" (HCV) refers to an agent causative of Non-A, Non-B Hepatitis (NANBH). The nucleic acid sequence and putative amino acid sequence of HCV is described in U.S. Patent Nos. 5,856,437 and 5,350,671. The disease caused by HCV is called hepatitis C, formerly called NANBH. The term HCV, as used herein, denotes a viral species of which pathenogenic strains cause NANBH, as well as attenuated strains or defective interfering particles derived therefrom.

HCV is a member of the viral family flaviviridae. The morphology and composition of Flavivirus particles are known, and are discussed in Reed et al., *Curr. Stud.-Hematol. Blood Transfus.* (1998), 62:1-37; HEPATITIS C VIRUSES IN FIELDS VIROLOGY (B.N. Fields, D.M. Knipe, P.M. Howley, eds.) (3d ed. 1996). It has recently been found that portions of the HCV genome are also homologous to pestiviruses. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

The HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation. Therefore, there can be

multiple strains, which can be virulent or avirulent, within the HCV class or species. The ORF of HCV, including the translation spans of the core, non-structural, and envelope proteins, is shown in U.S. Patent Nos. 5,856,437 and 5,350,671.

5

10

15

20

25

30

The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, multimers, and the like, are included within the definition. Both full-length proteins and fragments thereof are encompassed by the definition. The terms also include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation and the like. Furthermore, for purposes of the present invention, a "polypeptide" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the desired activity. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the proteins or errors due to PCR amplification.

An HCV polypeptide is a polypeptide, as defined above, derived from the HCV polyprotein. The polypeptide need not be physically derived from HCV, but may be synthetically or recombinantly produced. Moreover, the polypeptide may be derived from any of the various HCV strains, such as from strains 1, 2, 3 or 4 of HCV. A number of conserved and variable regions are known between these strains and, in general, the amino acid sequences of epitopes derived from these regions will have a high degree of sequence homology, e.g., amino acid sequence homology of more than 30%, preferably more than 40%, when the two sequences are aligned and homology determined by any of the programs or algorithms described herein. Thus, for example, the term "NS4" polypeptide refers to native NS4 from any of the various HCV strains, as well as NS4 analogs, muteins and immunogenic fragments, as defined further below.

Further, the terms "ΔNS35," "delNS35," "ΔNS3NS5," and "ΔNS3-5" as used herein refer to a mutant polypeptide, comprising at least portions of NS3, NS4, or NS5, comprising a deletion in, or mutation of, the NS3 protease active site region to render the protease non-functional. In one embodiment, ΔNS3-5 comprises amino acids 1242-3011, as shown in FIG. 5, or polypeptides substantially homologous thereto. It will be readily apparent to one of ordinary skill in the art how to determine that NS3 protease

has been rendered non-functional. If the protease is functional, one will obtain protein of the expected molecular weight upon expression. As set forth in Example 2 and Figure 15, using SDS-page, 4-20%, a protein having a molecular weight of approximately 194kD was obtained when strain AD3 was transformed with pd.ΔNS3NS5.PJ clone #5. One skilled in the art could readily determine whether a protein of the desired molecular weight was expressed for any given deletion or mutation.

5

10

15

20

25

30

The terms "analog" and "mutein" refer to biologically active derivatives of the reference molecule, or fragments of such derivatives, that retain desired activity, such as the ability to stimulate a cell-mediated immune response, as defined below. In general, the term "analog" refers to compounds having a native polypeptide sequence and structure with one or more amino acid additions, substitutions (generally conservative in nature) and/or deletions, relative to the native molecule, so long as the modifications do not destroy immunogenic activity. The term "mutein" refers to peptides having one or more peptide mimics ("peptoids"), such as those described in International Publication No. WO 91/04282. Preferably, the analog or mutein has at least the same immunoactivity as the native molecule. Methods for making polypeptide analogs and muteins are known in the art and are described further below.

Particularly preferred analogs include substitutions that are conservative in nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cysteine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. For example, it is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the biological activity. For example, the polypeptide of interest may include up to about 5-10 conservative or non-conservative amino acid substitutions, or even up to about 15-25 conservative or non-conservative amino acid substitutions, or any integer between 5-25, so long as the

desired function of the molecule remains intact. One of skill in the art may readily determine regions of the molecule of interest that can tolerate change by reference to Hopp/Woods and Kyte-Doolittle plots, well known in the art.

5

10

15

20

25

30

By "fragment" is intended a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure. The fragment can include a C-terminal deletion and/or an N-terminal deletion of the native polypeptide. An "immunogenic fragment" of a particular HCV protein will generally include at least about 5-10 contiguous amino acid residues of the full-length molecule, preferably at least about 15-25 contiguous amino acid residues of the full-length molecule, and most preferably at least about 20-50 or more contiguous amino acid residues of the full-length molecule, that define an epitope, or any integer between 5 amino acids and the full-length sequence, provided that the fragment in question retains immunogenic activity, as measured by the assays described herein. For a description of various HCV epitopes, see, e.g., Chien et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:10011-10015; Chien et al., *J. Gastroent. Hepatol.* (1993) 8:S33-39; Chien et al., International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; commonly owned, allowed U.S. Patent Application Serial Nos. 08/403,590 and 08/444,818.

The term "epitope" as used herein refers to a sequence of at least about 3 to 5, preferably about 5 to 10 or 15, and not more than about 1,000 amino acids (or any integer therebetween); which define a sequence that by itself or as part of a larger sequence, binds to an antibody generated in response to such sequence. There is no critical upper limit to the length of the fragment, which may comprise nearly the full-length of the protein sequence, or even a fusion protein comprising two or more epitopes from the HCV polyprotein. An epitope for use in the subject invention is not limited to a polypeptide having the exact sequence of the portion of the parent protein from which it is derived. Indeed, viral genomes are in a state of constant flux and contain several variable domains which exhibit relatively high degrees of variability between isolates. Thus the term "epitope" encompasses sequences identical to the native sequence, as well as modifications to the native sequence, such as deletions, additions and substitutions (generally conservative in nature).

5

10

15

20

25

30

Regions of a given polypeptide that include an epitope can be identified using any number of epitope mapping techniques, well known in the art. See, e.g., Epitope Mapping Protocols in Methods in Molecular Biology, Vol. 66 (Glenn E. Morris, Ed., 1996) Humana Press, Totowa, New Jersey. For example, linear epitopes may be determined by e.g., concurrently synthesizing large numbers of peptides on solid supports, the peptides corresponding to portions of the protein molecule, and reacting the peptides with antibodies while the peptides are still attached to the supports. Such techniques are known in the art and described in, e.g., U.S. Patent No. 4,708,871; Geysen et al. (1984) Proc. Natl. Acad. Sci. USA 81:3998-4002; Geysen et al. (1986) Molec. Immunol. 23:709-715. Similarly, conformational epitopes are readily identified by determining spatial conformation of amino acids such as by, e.g., x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., Epitope Mapping Protocols, supra. Antigenic regions of proteins can also be identified using standard antigenicity and hydropathy plots, such as those calculated using, e.g., the Omiga version 1.0 software program available from the Oxford Molecular Group. This computer program employs the Hopp/Woods method, Hopp et al., Proc. Natl. Acad. Sci USA (1981) 78:3824-3828 for determining antigenicity profiles, and the Kyte-Doolittle technique, Kyte et al., J. Mol. Biol. (1982) 157:105-132 for hydropathy plots.

As used herein, the term "conformational epitope" refers to a portion of a full-length protein, or an analog or mutein thereof, having structural features native to the amino acid sequence encoding the epitope within the full-length natural protein. Native structural features include, but are not limited to, glycosylation and three dimensional structure. Preferably, a conformational epitope is produced recombinantly and is expressed in a cell from which it is extractable under conditions which preserve its desired structural features, e.g. without denaturation of the epitope. Such cells include bacteria, yeast, insect, and mammalian cells. Expression and isolation of recombinant conformational epitopes from the HCV polyprotein are described in e.g., International Publication Nos. WO 96/04301, WO 94/01778, WO 95/33053, WO 92/08734.

جوجو ال

An "immunological response" to an HCV antigen (including both polypeptide and polynucleotides encoding polypeptides that are expressed *in vivo*) or composition is the development in a subject of a humoral and/or a cellular immune response to molecules present in the composition of interest. For purposes of the present invention,

a "humoral immune response" refers to an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTLs"). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the intracellular destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

5

10

15

20

25

30

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376; and the examples below.

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate

antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection or alleviation of symptoms to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

5

10

15

20

25

30

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A transcription termination sequence may be located 3' to the coding sequence.

A "nucleic acid" molecule or "polynucleotide" can include both double- and single-stranded sequences and refers to, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral (e.g. DNA viruses and retroviruses) or procaryotic DNA, and especially synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their desired function. Thus, a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper transcription factors, etc., are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence, as can transcribed introns, and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, viral, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation is not associated with all or a portion of the polynucleotide with which it is associated in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. In general, the gene of interest is cloned and then

expressed in transformed organisms, as described further below. The host organism expresses the foreign gene to produce the protein under expression conditions.

A "control element" refers to a polynucleotide sequence which aids in the expression of a coding sequence to which it is linked. The term includes promoters, transcription termination sequences, upstream regulatory domains, polyadenylation signals, untranslated regions, including 5'-UTRs and 3'-UTRs and when appropriate, leader sequences and enhancers, which collectively provide for the transcription and translation of a coding sequence in a host cell.

5

10

15

20

25

30

A "promoter" as used herein is a DNA regulatory region capable of binding RNA polymerase in a host cell and initiating transcription of a downstream (3' direction) coding sequence operably linked thereto. For purposes of the present invention, a promoter sequence includes the minimum number of bases or elements necessary to initiate transcription of a gene of interest at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eucaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

A control sequence "directs the transcription" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

"Expression cassette" or "expression construct" refers to an assembly which is capable of directing the expression of the sequence(s) or gene(s) of interest. The expression cassette includes control elements, as described above, such as a promoter which is operably linked to (so as to direct transcription of) the sequence(s) or gene(s) of interest, and often includes a polyadenylation sequence as well. Within certain embodiments of the invention, the expression cassette described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), at least one multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for insertion: for example, transformation by direct uptake, transfection, infection, and the like. For particular methods of transfection, see further below. The exogenous polynucleotide may be maintained as a nonintegrated vector, for example, an episome, or alternatively, may be integrated into the host genome.

A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous DNA sequence.

5

10

15

20

25

30

By "isolated" is meant, when referring to a polypeptide, that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macromolecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in association therewith; or a molecule disassociated from the chromosome.

The term "purified" as used herein preferably means at least 75% by weight, more preferably at least 85% by weight, more preferably still at least 95% by weight, and most preferably at least 98% by weight, of biological macromolecules of the same type are present.

"Homology" refers to the percent identity between two polynucleotide or two polypeptide moieties. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 50%, preferably at least about 75%, more preferably at least about 80%-85%, preferably at least about 90%, and most preferably at least about 95%-98%, or more, sequence identity over a defined length of the molecules. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. The term "substantially homologous" as used herein in reference to ΔNS35 generally refers to an HCV nucleic or amino acid sequence that is at least 60% identical to the entire sequence of the polypeptide encoded by ΔNS35 (see FIG. 5), where the sequence identity is preferably at least 75%, more preferably at least 80%, still more preferably at least about 85%, especially more than about 90%, most preferably 95% or greater, particularly 98% or greater. These homologous polypeptides include fragments,

including mutants and allelic variants of the fragments. Identity between the two sequences is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters gap open penalty=12 and gap extension penalty=1. Thus, for example, the present invention includes an isolate which is 80% identical to a polypeptide encoded by Δ NS35. In some aspects of the invention, the polypeptide of the present invention is substantially homologous to the Δ NS35.

5

10

15

20

25

30

In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acidto-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Percent identity can be determined by a direct comparison of the sequence information between two molecules by aligning the sequences, counting the exact number of matches between the two aligned sequences, dividing by the length of the shorter sequence, and multiplying the result by 100. Readily available computer programs can be used to aid in the analysis, such as ALIGN, Dayhoff, M.O. in Atlas of Protein Sequence and Structure M.O. Dayhoff ed., 5 Suppl. 3:353-358, National biomedical Research Foundation, Washington, DC, which adapts the local homology algorithm of Smith and Waterman Advances in Appl. Math. 2:482-489, 1981 for peptide analysis. Programs for determining nucleotide sequence identity are available in the Wisconsin Sequence Analysis Package, Version 8 (available from Genetics Computer Group, Madison, WI) for example, the BESTFIT, FASTA and GAP programs, which also rely on the Smith and Waterman algorithm. These programs are readily utilized with the default parameters recommended by the manufacturer and described in the Wisconsin Sequence Analysis Package referred to above. For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions.

Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a

200 200

gap of six). From the data generated the "Match" value reflects "sequence identity."

Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address:

http://www.ncbi.nlm.gov/cgi-bin/BLAST.

5

10

15

20

25

30

Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; *DNA Cloning*, *supra*; *Nucleic Acid Hybridization*, *supra*.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1µg for a plasmid or phage digest to 10° to 10° g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a

smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10⁸ cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10⁸ cpm/µg, resulting in an exposure time of ~24 hours.

5

10

15

20

25

30

F 5120

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

Tm= $81 + 16.6(\log_{10}\text{Ci}) + 0.4[\%(G + C)]-0.6(\%\text{formamide}) - 600/n-1.5(\%\text{mismatch}).$ where Ci is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) Anal. Biochem. 138: 267-284). In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of the antigen or antigens. The nucleic acid molecule can be introduced directly into the recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have

been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

An "open reading frame" or ORF is a region of a polynucleotide sequence which encodes a polypeptide; this region can represent a portion of a coding sequence or a total coding sequence.

5

10

15

20

25

30

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which comprise at least one antigen binding site. An "antigen binding site" is formed from the folding of the variable domains of an antibody molecule(s) to form three-dimensional binding sites with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows specific binding to form an antibody-antigen complex. An antigen binding site may be formed from a heavy- and/or light-chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, without limitation, polyclonal antibodies, monoclonal antibodies, chimeric antibodies, altered antibodies, univalent antibodies, Fab proteins, and single-domain antibodies. In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker, eds. (1987) IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).

Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g.,

M. Schreier et al. (1980) HYBRIDOMA TECHNIQUES; Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS; Kennett et al. (1980) MONOCLONAL ANTIBODIES; see also, U.S. Pat. Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc. As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an HL domain, which binds specifically with a designated antigen. A dAb does not contain a VL domain, but may contain other antigen binding domains known to exist to antibodies, for example, the kappa and lambda domains. Methods for preparing dabs are known in the art. See, for example, Ward et al, Nature 341: 544 (1989).

5

10

15

20

25

30

رية وي

Antibodies can also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation and known in the art (see, e.g., U.S. Pat. No. 4,816,467), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of the chains are homologous with those sequences found in antibodies produced in vertebrates, whether in situ or in vitro (for example, in hybridomas). Vertebrate antibodies include, for example, purified polyclonal antibodies and monoclonal antibodies, methods for the preparation of which are described infra.

"Hybrid antibodies" are antibodies where chains are separately homologous with reference to mammalian antibody chains and represent novel assemblies of them, so that two different antigens are precipitable by the tetramer or aggregate. In hybrid antibodies, one pair of heavy and light chains are homologous to those found in an antibody raised against a first antigen, while a second pair of chains are homologous to those found in an antibody raised against a second antibody. This results in the property of "divalence", i.e., the ability to bind two antigens simultaneously. Such hybrids can also be formed using chimeric chains, as set forth below.

"Chimeric antibodies" refers to antibodies in which the heavy and/or light chains are fusion proteins. Typically, one portion of the amino acid sequences of the

5

10

15

20

25

30

chain is homologous to corresponding sequences in an antibody derived from a particular species or a particular class, while the remaining segment of the chain is homologous to the sequences derived from another species and/or class. Usually, the variable region of both light and heavy chains mimics the variable regions or antibodies derived from one species of vertebrates, while the constant portions are homologous to the sequences in the antibodies derived from another species of vertebrates. However, the definition is not limited to this particular example. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be from differing classes or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic know antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varies. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region can be made to alter antigen binding characteristics. The antibody can also be engineered to aid the specific delivery of a molecule or substance to a specific cell or tissue site. The desired alterations can be made by known techniques in molecular biology, e.g., recombinant techniques, site-directed mutagenesis, etc.

ه . اين

Yet another example are "univalent antibodies", which are aggregates comprised of a heavy-chain/light-chain dimer bound to the Fc (i.e., stem) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. Nature 295: 712 (1982). Included also within the definition of antibodies

are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion. "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as F(ab)2), which are capable of selectively reacting with a designated antigen or antigen family. Fab antibodies can be divided into subsets analogous to those described above, i.e., "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing Fab fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

5

10

15

20

25

30

"Antigen-antibody complex" refers to the complex formed by an antibody that is specifically bound to an epitope on an antigen.

"Immunogenic polypeptide" refers to a polypeptide that elicits a cellular and/or humoral immune response in a mammal, whether alone or linked to a carrier, in the presence or absence of an adjuvant.

"Antigenic determinant" refers to the site on an antigen or hapten to which a specific antibody molecule or specific cell surface receptor binds.

As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

By "vertebrate subject" is meant any member of the subphylum cordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The

invention described herein is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

II. Modes of Carrying out the Invention

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of compositions and methods similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

General Overview

5

10

15

20

25

30

An aim of an HCV vaccine is to generate broad immunity to a wide breadth of antigens because HCV is so divergent and because humoral as well as cellular immune responses are desirable to combat this human pathogen. While antibodies generated against the envelope glycoprotein(s) might aid in virus neutralization, there is additional benefit to be derived from a vaccine that includes other regions. The likelihood of T-helper responses generated against a polypeptide would be helpful in a vaccine setting as would generation of cytotoxic T cells. The non-structural region represents such a candidate antigen, but processing by the protease generates several polypeptides, making purification complicated. It would be advantageous, therefore, to derive a non-structural cassette that is unprocessed by the NS3 protease.

The present invention solves this and other problems using compositions and methods involving an N-terminal deletion in NS3, which removes the catalytic domain. As such, some or all of the remainder of the non-structural region (through NS5B) is expressed as an intact polypeptide. Expression of this species has been documented in mammalian cells as well as in yeast. Further, in certain aspects, polynucleotides encoding HCV core polypeptides (or fragments thereof) are added (e.g., operably linked) to the carboxy-terminus of the non-structural cassette. As the core coding region is relatively highly conserved among HCV isolates, the presence of this region

may enhance the immune response. Because core has at its C-terminus a very hydrophobic domain (amino acids 174-191), shorter versions of core were also engineered onto the polypeptide. As described in detail herein, the truncation of core to amino acid 121 yielded higher expression than the amino acid 173 truncation when engineered onto the C-terminus of the mutant NS polypeptide. The combination of most of the non-structural region fused to a C-terminally truncated core into a polypeptide is novel and has advantages for vaccine immunization. Moreover, because the aim is not necessarily to generate antibody responses to this polypeptide, there is no need to maintain a native conformation, enabling a more facile purification protocol.

10

15

20

25

5

Mutant HCV Non-Structural Polypeptides

Genomes of HCV strains contain a single open reading frame of approximately 9,000 to 12,000 nucleotides, which is transcribed into a polyprotein. An HCV polyprotein is cleaved to produce at least ten distinct products, in the order of NH₂-Core-E1-E2-p7-NS2-NS3-NS4a-NS4b-NS5a-NS5b-COOH. Mutant HCV polypeptides of the invention contain an N-terminal deletion in NS3, which removes or disables the catalytic domain. Preferably, the polypeptides also include the remainder of the non-structural region, although in certain embodiments, the polypeptides may include less than all of the remaining NS polypeptides, for example mutant NS polypeptides including any combinations of NS2-NS3-NS4a-NS4b-NS5a-NS5b (e.g., NS3NS3-NS5a-NS5b; NS3-NS4a-NS4b-NS5a; NS3-NS4b-NS5a; NS3-NS4b-NS5a; NS3-NS4b-NS5b; etc.).

The HCV-NS3 protein functions as a protease and a helicase and occurs at approximately amino acid 1027 to amino acid 1657 of the polyprotein (numbered relative to HCV-1). See Choo et al. (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455. HCV NS4 occurs at approximately amino acid 1658 to amino acid 1972, NS5a occurs at approximately amino acid 1973 to amino acid 2420, and HCV NS5b occurs at approximately amino acid 2421 to amino acid 3011 of the polyprotein (numbered relative to HCV-1) (Choo et al., 1991).

30

The mutant polypeptides described herein can either be full-length polypeptides or portions of NS3, NS4 (NS4a and NS4b), NS5a, and NS5b polypeptides. Epitopes of NS3, NS4 (NS4a and NS4b), NS5a, NS5b, NS3NS4NS5a, and NS3NS4NS5aNS5b can

;

be identified by several methods. For example, NS3, NS4, NS5a, NS5b polypeptides or fusion proteins comprising any combination of the above, can be isolated, for example, by immunoaffinity purification using a monoclonal antibody for the polypeptide or protein. The isolated protein sequence can then be screened by preparing a series of short peptides by proteolytic cleavage of the purified protein, which together span the entire protein sequence. By starting with, for example, 100-mer polypeptides, each polypeptide can be tested for the presence of epitopes recognized by a T cell receptor on an HCV-activated T cell, progressively smaller and overlapping fragments can then be tested from an identified 100-mer to map the epitope of interest.

5

10

15

20

25

30

Epitopes recognized by a T cell receptor on an HCV-activated T cell can be identified by, for example, ⁵¹Cr release assay (see Example 2) or by lymphoproliferation assay (see Example 4). In a ⁵¹Cr release assay, target cells can be constructed that display the epitope of interest by cloning a polynucleotide encoding the epitope into an expression vector and transforming the expression vector into the target cells. Non-structural polypeptides can occur in any order in the fusion protein. If desired, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more of one or more of the polypeptides may occur in the fusion protein. Multiple viral strains of HCV occur, and NS3, NS4, NS5a, and NS5b polypeptides of any of these strains can be used in a fusion protein.

Nucleic acid and amino acid sequences of a number of HCV strains and isolates, including nucleic acid and amino acid sequences of NS3, NS4, NS5a, NS5b genes and polypeptides have been determined. For example, isolate HCV J1.1 is described in Kubo et al. (1989) Japan. Nucl. Acids Res. 17:10367-10372; Takeuchi et al. (1990) Gene 91:287-291; Takeuchi et al. (1990) J. Gen. Virol. 71:3027-3033; and Takeuchi et al. (1990) Nucl. Acids Res. 18:4626. The complete coding sequences of two independent isolates, HCV-J and BK, are described by Kato et al., (1990) Proc. Natl. Acad. Sci. USA 87:9524-9528 and Takamizawa et al., (1991) J. Virol. 65:1105-1113 respectively.

Publications that describe HCV-1 isolates include Choo et al. (1990) Brit. Med. Bull. 46:423-441; Choo et al. (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455 and Han et al. (1991) Proc. Natl. Acad. Sci. USA 88:1711-1715. HCV isolates HC-J1 and HC-J4 are described in Okamoto et al. (1991) Japan J. Exp. Med. 60:167-177. HCV

isolates HCT 18~, HCT 23, Th, HCT 27, EC1 and EC10 are described in Weiner et al. (1991) Virol. 180:842-848. HCV isolates Pt-1, HCV-K1 and HCV-K2 are described in Enomoto et al. (1990) Biochem. Biophys. Res. Commun. 170:1021-1025. HCV isolates A, C, D & E are described in Tsukiyama-Kohara et al. (1991) Virus Genes 5:243-254.

5

10

15

20

+5 4 ft

25

30

Each of the mutant HCV polypeptides containing at least portions of NS3, NS4 and NS5 can be obtained from the same HCV strain or isolate or from different HCV strains or isolates. Thus, each non-structural region of the polypeptide can be from the same HCV strain or isolate or from each different HCV strains or isolates. In addition to the mutant HCV non-structural polypeptides described herein, the proteins can contain other polypeptides derived from the HCV polyprotein. For example, it may be desirable to include polypeptides derived from the core region of the HCV polyprotein. This region occurs at amino acid positions 1-191 of the HCV polyprotein, numbered relative to HCV-1. Either the full-length protein or epitopes of the full-length protein may be used in the subject fusions, such as those epitopes found between amino acids 10-53, amino acids 10-45, amino acids 67-88, amino acids 120-130, or any of the core epitopes identified in, e.g., Houghton et al., U.S. Patent No. 5,350,671; Chien et al., Proc. Natl. Acad. Sci. USA (1992) 89:10011-10015; Chien et al., J. Gastroent, Hepatol. (1993) 8:S33-39; Chien et al., International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; and commonly owned, U.S. Patent No. 6,150,087. When present, additional non-structural HCV polypeptides such as core can be obtained from the same HCV strain or isolate or from different HCV strains or isolates.

Preferably, the above-described mutant proteins, as well as the individual components of these proteins, are produced recombinantly. A polynucleotide encoding these proteins can be introduced into an expression vector which can be expressed in a suitable expression system. A variety of bacterial, yeast, mammalian, insect and plant expression systems are available in the art and any such expression system can be used. Optionally, a polynucleotide encoding these proteins can be translated in a cell-free translation system. Such methods are well known in the art. The proteins also can be constructed by solid phase protein synthesis.

If desired, the mutant polypeptides, or the individual components of these polypeptides, also can contain other amino acid sequences, such as amino acid linkers or signal sequences, as well as ligands useful in protein purification, such as glutathione-S-transferase and staphylococcal protein A.

5

10

15

Polynucleotides

The polynucleotides of the present invention are not necessarily physically derived from the nucleotide sequences shown, but can be generated in any manner, including, for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequences can be modified in ways known to the art to be consistent with an intended use.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, can be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell is given below. The polypeptide produced in such host cells is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use.

20

Purification can be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, alkali resolubilization of insoluble protein, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

25

Polynucleotides contain less than an entire HCV genome and can be RNA or single- or double-stranded DNA. Preferably, the polynucleotides are isolated free of other components, such as proteins and lipids. Polynucleotides of the invention can also comprise other nucleotide sequences, such as sequences coding for linkers, signal sequences, or ligands useful in protein purification such as glutathione-S-transferase and staphylococcal protein A.

30

Polynucleotides encoding mutant HCV non-structural polypeptides can be isolated from a genomic library derived from nucleic acid sequences present in, for

1

example, the plasma, serum, or liver homogenate of an HCV infected individual or can be synthesized in the laboratory, for example, using an automatic synthesizer. An amplification method such as PCR can be used to amplify polynucleotides from either HCV genomic DNA or cDNA.

5

10

15

20

25

30

Further, while the polypeptides that are not NS3, NS4, or NS5 of HCV of the present invention can comprise a substantially complete viral domain, in many applications all that is required is that the polypeptide comprise an antigenic or immunogenic region of the virus. An antigenic region of a polypeptide is generally relatively small-typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids can characterize an antigenic region. These segments can correspond to regions of, for example, C, E1, or E2 epitopes. Accordingly, using the cDNAs of C, E1, or E2 as a basis, DNAs encoding short segments of C, E1, or E2 polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis.

Polynucleotides encoding the polypeptides described herein can comprise coding sequences for these polypeptides which occur naturally or can be artificial sequences which do not occur in nature. These polynucleotides can be ligated to form a coding sequence for the fusion proteins using standard molecular biology techniques. If desired, polynucleotides can be cloned into an expression vector and transformed into, for example, bacterial, yeast, insect, plant or mammalian cells so that the fusion proteins of the invention can be expressed in and isolated from a cell culture.

The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, plant and vertebrate cells, give rise to important immunological reagents which can be used for diagnosis, detection, and vaccines.

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and. ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general

guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

5

10

15

20

25

30

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, E. coli is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977), Nature 198:1056), the tryptophan (trp) promoter system (Goeddel et al. (1980) Nucleic Acid Res. 8:4057), the lambda-derived P[L | promoter and N gene ribosome binding site (Shimatake et al. (1981) Nature 292:128) and the hybrid tac promoter (De Boer et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 292:128) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with E. coli; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include mammalian and yeast cells in culture systems. Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978), Nature 273:113), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or

sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

The vaccinia virus system can also be used to express foreign DNA in mammalian cells. To express heterologous genes, the foreign DNA is usually inserted into the thymidine kinase gene of the vaccinia virus and then infected cells can be selected. This procedure is known in the art and further information can be found in these references (Mackett et al. J. Virol. 49: 857-864 (1984) and Chapter 7 in DNA Cloning, Vol. 2, IRL Press).

5

10

15

20

25

30

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences (Myanohara et al. (1983) *Proc. Natl. Acad. Sci. USA 80*:1).

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197).

and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the ADH2, GAL4, GAL10, OR PHO5 genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, inter alia, (Cohen et al. (1980) Proc. Natl. Acad. Sci. USA 77:1078; Henikoff et al. (1981) Nature 283:835; Hollenberg et al. (1981) Curr. Topics Microbiol. Immunol. 96:119; Hollenberg et al. (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: Plasmids of Medical, Environmental and Commercial Importance (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon et al. (1980) Gene 11:163; Panthier et al. (1980) Curr. Genet. 2:109).

5

10

15

20

25

30

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See e.g., EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (e.g., ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (e.g., WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a

leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

5

10

15

20

25

30

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (e.g., see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle

vectors include YEp24 (Botstein et al. (1979) Gene 8:17-24), pCl/1 (Brake et al. (1984) Proc. Natl. Acad. Sci USA 81:4642-4646), and YRp17 (Stinchcomb et al. (1982) J. Mol. Biol. 158:157). In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See e.g., Brake et al., supra.

5

10

15

20

25

30

is em

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome (Orr-Weaver et al. (1983) Methods in Enzymol. 101:228-245). An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver et al., supra. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced (Rine et al. (1983) Proc. Natl. Acad. Sci. USA 80:6750). The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as ADE2, HIS4, LEU2, TRP1, and ALG7, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of CUP1

allows yeast to grow in the presence of copper ions (Butt et al. (1987) Microbiol, Rev. 51:351).

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

5

25

30

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, inter alia, the following yeasts: Candida albicans (Kurtz, et al. (1986) Mol. Cell. Biol. 6:142), Candida maltosa 10 (Kunze, et al. (1985) J. Basic Microbiol. 25:141). Hansenula polymorpha (Gleeson, et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302), Kluyveromyces fragilis (Das, et al. (1984) J. Bacteriol. 158:1165), Kluyveromyces lactis (De Louvencourt et al. (1983) J. Bacteriol. 154:737; Van den 15 Berg et al. (1990) Bio/Technology 8:135), Pichia guillerimondii (Kunze et al. (1985) J. Basic Microbiol. 25:141), Pichia pastoris (Cregg, et al. (1985) Mol. Cell. Biol. 5:3376; US Patent Nos. 4,837,148 and 4,929,555), Saccharomyces cerevisiae (Hinnen et al. (1978) Proc. Natl. Acad. Sci. USA 75:1929; Ito et al. (1983) J. Bacteriol. 153:163), Schizosaccharomyces pombe (Beach and Nurse (1981) Nature 300:706), and 20 Yarrowia lipolytica (Davidow, et al. (1985) Curr. Genet. 10:380471 Gaillardin, et al. (1985) Curr. Genet. 10:49).

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. (See e.g., Kurtz et al. (1986) Mol. Cell. Biol. 6:142; Kunze et al. (1985) J. Basic Microbiol. 25:141; Candida; Gleeson et al. (1986) J. Gen. Microbiol. 132:3459; Roggenkamp et al. (1986) Mol. Gen. Genet. 202:302; Hansenula; Das et al. (1984) J. Bacteriol. 158:1165; De Louvencourt et al. (1983) J. Bacteriol. 154:1165; Van den Berg et al. (1990) Bio/Technology 8:135; Kluyveromyces; Cregg et al. (1985) Mol. Cell. Biol. 5:3376; Kunze et al. (1985) J. Basic Microbiol. 25:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia; Hinnen et

al. (1978) Proc. Natl. Acad. Sci. USA 75;1929; Ito et al. (1983) J. Bacteriol.

1

153:163 Saccharomyces; Beach and Nurse (1981) Nature 300:706; Schizosaccharomyces; Davidow et al. (1985) Curr. Genet. 10:39; Gaillardin et al. (1985) Curr. Genet. 10:49; Yarrowia).

5

10

15

20

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E. coli) (Raibaud et al. (1984) Annu. Rev. Genet. 18:173). Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, inter alia, the following bacteria:

Bacillus subtilis (Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541), Escherichia coli (Shimatake et al. (1981) Nature 292:128; Amann et al. (1985) Gene 40:183; Studier et al. (1986) J. Mol. Biol. 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907), Streptococcus cremoris (Powell et al. (1988) Appl. Environ. Microbiol. 54:655); Streptococcus lividans (Powell et al. (1988) Appl. Environ. Microbiol. 54:655), Streptomyces lividans (US patent 4,745,056).

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl, or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. (See e.g., Masson et al. (1989) FEMS Microbiol. 5 Lett. 60:273; Palva et al. (1982) Proc. Natl. Acad. Sci. USA 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus, Miller et al. (1988) Proc. Natl. Acad. Sci. 85:856; Wang et al. (1990) J. Bacteriol. 172:949; Campylobacter, Cohen et al. (1973) Proc. Natl. Acad. Sci. 69:2110; Dower et al. (1988) Nucleic Acids Res. 10 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering (eds. H.W. Boyer and S. Nicosia): Mandel et al. (1970) J. Mol. Biol. 53:159; Taketo (1988) Biochim. Biophys. Acta 949:318; Escherichia; Chassy et al. (1987) FEMS Microbiol. Lett. 44:173 15 Lactobacillus; Fiedler et al. (1988) Anal. Biochem 170:38, Pseudomonas; Augustin et al. (1990) FEMS Microbiol. Lett. 66:203, Staphylococcus, Barany et al. (1980) J. Bacteriol. 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: Streptococcal Genetics (ed. J. Ferretti and R. Curtiss III); Perry et al. (1981) Infect. Immun. 32:1295; Powell et al. (1988) Appl. Environ. Microbiol. 20 54:655; Somkuti et al. (1987) Proc. 4th Evr. Cong. Biotechnology 1:412, Streptococcus).

In addition, viral antigens can be expressed in insect cells by the Baculovirus system. A general guide to Baculovirus expression by Summer and Smith is A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures (Texas

25 Agricultural Experiment Station Bulletin No. 1555). To incorporate the heterologous gene into the Baculovirus genome the gene is first cloned into a transfer vector containing some Baculovirus sequences. This transfer vector, when it is cotransfected with wild-type virus into insect cells, will recombine with the wild-type virus. Usually, the transfer vector will be engineered so that the heterologous gene will disrupt the

30 wild-type Baculovirus polyhedron gene. This disruption enables easy selection of the recombinant virus since the cells infected with the recombinant virus will appear phenotypically different from the cells infected with the wild-type virus. The purified

recombinant virus can be used to infect cells to express the heterologous gene. The foreign protein can be secreted into the medium if a signal peptide is linked in frame to the heterologous gene; otherwise, the protein will be bound in the cell lysates. For further information, see Smith et al Mol. & Cell. Biol. 3:2156-2165 (1983) or Luckow and Summers in Virology 17: 31-39 (1989).

5

10

15

20

25

30

Baculovirus expression can also be affected in plant cells. There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, Phytochemistry 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., Mol. Gen. Genet. 209:33-40 (1987); Chandler et al., Plant Molecular Biology 3:407-418 (1984); Rogers, J. Biol. Chem. 260:3731-3738 (1985); Rothstein et al., Gene 55:353-356 (1987); Whittier et al., Nucleic Acids Research 15:2515-2535 (1987); Wirsel et al., Molecular Microbiology 3:3-14 (1989); Yu et al., Gene 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: Advanced Plant Physiology, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolicallyregulated genes: Sheen, Plant Cell, 2:1027-1038(1990); Maas et al., EMBO J. 9:3447-3452 (1990); Benkel and Hickey, Proc. Natl. Acad. Sci. 84:1337-1339 (1987).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersion, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis,

Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

5

10

15

20

25

30

Transformation can be by any method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972), Proc. Natl. Acad. Sci. U.S.A. 69:2110; Maniatis et al. (1982), MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978) Proc. Natl. Acad. Sci. U.S.A. 75: 1929. Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), Virology 52:546 or the various known modifications thereof.

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560. Sticky ended cleavage fragments may be blunt ended using E. coli DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation. Ligation mixtures are transformed into suitable cloning hosts,

such as E. coli, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

5

10

15

25

30

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984), DNA 3:401. If desired, the synthetic strands may be labeled with 32P by treatment with polynucleotide kinase in the presence of 32P-ATP, using standard conditions for the reaction. DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982), Nucleic Acids Res. 10:6487.

The expression constructs of the present invention, including the desired fusion, or individual expression constructs comprising the individual components of these fusions, may be used for nucleic acid immunization, to activate HCV-specific T cells, using standard gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered ex vivo, to cells derived from the subject and the cells reimplanted in the subject. For example, the constructs can be delivered as plasmid DNA, e.g., contained within a plasmid, such as pBR322, pUC, or ColE1

Additionally, the expression constructs can be packaged in liposomes prior to 20 delivery to the cells. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, Biochim. Biophys. Acta. (1991) 1097:1-17; Straubinger et al., in Methods of Enzymology (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use with the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., Proc. Natl. Acad. Sci. USA (1987) 84:7413-7416). Other

commercially available lipids include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., Proc. Natl. Acad. Sci. USA (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a 5 description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes. The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al., Proc. Natl. Acad. Sci. USA (1978) 75:4194-4198; Papahadjopoulos et al., Biochim. Biophys. 10 Acta (1975) 394:483; Wilson et al., Cell (1979) 17:77); Deamer and Bangham, Biochim. Biophys. Acta (1976) 443:629; Ostro et al., Biochem. Biophys. Res. Commun. (1977) 76:836; Fraley et al., Proc. Natl. Acad. Sci. USA (1979) 76:3348); Enoch and Strittmatter, Proc. Natl. Acad. Sci. USA (1979) 76:145); Fraley et al., J. Biol. Chem. (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., Science (1982) 215:166. 15

The DNA can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochem. Biophys. Acta.* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

20

25

30

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems, such as murine sarcoma virus, mouse mammary tumor virus, Moloney murine leukemia virus, and Rous sarcoma virus. A selected gene can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109. Briefly, retroviral gene delivery vehicles of the present invention may be readily constructed from a wide variety of retroviruses, including for example, B, C, and D type retroviruses as well as spumaviruses and lentiviruses such as FIV, HIV, HIV-1, HIV-2 and SIV (see RNA

Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Such retroviruses may be readily obtained from depositories or collections such as the American Type Culture Collection ("ATCC"; 10801 University Blvd., Manassas, VA 20110-2209), or isolated from known sources using commonly available techniques.

5

10

15

20

25

30

A number of adenovirus vectors have also been described, such as adenovirus Type 2 and Type 5 vectors. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) <u>57</u>:267-274; Bett et al., *J. Virol.* (1993) <u>67</u>:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) <u>5</u>:717-729; Seth et al., *J. Virol.* (1994) <u>68</u>:933-940; Barr et al., *Gene Therapy* (1994) <u>1</u>:51-58; Berkner, K.L. *BioTechniques* (1988) <u>6</u>:616-629; and Rich et al., *Human Gene Therapy* (1993) <u>4</u>:461-476).

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as but not limited to vectors derived from the Sindbis and Semliki Forest viruses, VEE, will also find use as viral vectors for delivering the gene of interest. For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072.

Other vectors can be used, including but not limited to simian virus 40, cytomegalovirus. Bacterial vectors, such as Salmonella ssp. Yersinia enterocolitica, Shigella spp., Vibrio cholerae, Mycobacterium strain BCG, and Listeria monocytogenes can be used. Minichromosomes such as MC and MC1, bacteriophages, cosmids (plasmids into which phage lambda cos sites have been inserted) and replicons (genetic elements that are capable of replication under their own control in a cell) can also be used.

The expression constructs may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected molecule to the immune system and promote trapping and retention of molecules in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those

derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; and McGee et al., *J. Microencap.* (1996).

A wide variety of other methods can be used to deliver the expression constructs to cells. Such methods include DEAE dextran-mediated transfection, calcium phosphate precipitation, polylysine- or polyornithine-mediated transfection, or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like. Other useful methods of transfection include electroporation, sonoporation, protoplast fusion, liposomes, peptoid delivery, or microinjection. See, e.g., Sambrook et al., *supra*, for a discussion of techniques for transforming cells of interest; and Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. One particularly effective method of delivering DNA using electroporation is described in International Publication No. WO/0045823.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering the expression constructs of the present invention. The particles are coated with the construct to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744.

4400

Compositions

5

10

15

20

30

The invention also provides compositions comprising the HCV polypeptides or polynucleotides described herein. Such compositions are useful as diagnostics, for example, using the mutant polypeptides (or polynucleotides encoding these polypeptides) in diagnostic reagents. Diagnostics using polypeptides and polynucleotides are known to those of skill in the art.

In addition, immunogenic compounds can be prepared from one or more immunogenic polypeptides derived from the polypeptides described herein, for example the $\Delta NS35$ polypeptide. The preparation of immunogenic compounds which

contain immunogenic polypeptide(s) as active ingredients is known to one skilled in the art. Typically, such immunogenic compounds are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified, or the protein encapsulated in liposomes.

5

10

15

20

25

30

Immunogenic and diagnostic compositions of the invention preferably comprise a pharmaceutically acceptable carrier. The carrier should not itself induce the production of antibodies harmful to the host. Pharmaceutically acceptable carriers are well known to those in the art. Such carriers include, but are not limited to, large, slowly metabolized, macromolecules, such as proteins, polysaccharides such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like, polylactic acids, polyglycolic acids, polymeric amino acids such as polyglutamic acid, polylysine, and the like, amino acid copolymers, and inactive virus particles.

Pharmaceutically acceptable salts can also be used in compositions of the invention, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as salts of organic acids such as acetates, proprionates, malonates, or benzoates. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those of skill in the art. Compositions of the invention can also contain liquids or excipients, such as water, saline, glycerol, dextrose, ethanol, or the like, singly or in combination, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes can also be used as a carrier for a composition of the invention, such liposomes are described above.

If desired, co-stimulatory molecules which improve immunogen presentation to lymphocytes, such as B7-1 or B7-2, or cytokines such as GM-CSF, IL-2, and IL-12, can be included in a composition of the invention. Optionally, adjuvants can also be included in a composition. Adjuvants which can be used include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (PCT Publ. No. WO 90/14837).

containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE), formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer 5 L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, such as StimulonTM 10 (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g., gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor 15 (TNF), etc; (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an E. coli heat-labile toxin (LT), particularly LT-K63, LT-R72, CT-S109, PT-K9/G129; see, e.g., WO 93/13302 and WO 92/19265; (7) other substances that act as immunostimulating agents to enhance 20 the effectiveness of the composition; and (8) microparticles with adsorbed macromolecules, as described in copending U.S. Patent Application Serial No. 09/285,855 (filed April 2, 1999) and international Patent Application Serial No. PCT/US99/17308 (filed July 29, 1999). Alum and MF59 are preferred. The effectiveness of an adjuvant can be determined by measuring the amount of antibodies 25 directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in immunogenic compounds which are also comprised of the various adjuvants.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), -acetyl-normuramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), etc.

30

Thus, such recombinant or synthetic HCV polypeptides can be used in vaccines and as diagnostics. Further, antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, antibodies to these polypeptides are useful for isolating and identifying HCV particles.

Native HCV antigens can also be isolated from HCV virions. The virions can be grown in HCV infected cells in tissue culture, or in an infected host.

Administration and Delivery

5

10

15

20

25

30

The polynucleotide and polypeptide compositions described herein (e.g., immunogenic compounds) may be administered to a subject using any suitable delivery means. Methods of delivering nucleic acids into host cells are discussed above. Further, HCV polynucleotides and/or polypeptides can be administered parenterally, by injection, usually, subcutaneously, intramuscularly, transdermally or transcutaneously. Certain adjuvants, e.g. LTK63, LTR72 or PLG formulations, can be administered intranasally or orally. Additional formulations which are suitable for other modes of administration include suppositories. For suppositories, traditional binders and carriers can include, for example, polyalkylene glycols or triglycerides; such suppositories can be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Other oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The polypeptides of the present invention can be formulated into the immunogenic compound as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric

hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

5

10

15

20

25

30

The immunogenic compounds are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of polypeptide per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and can be peculiar to each subject.

The immunogenic compound can be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination can be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. Further, the course of administration may include polynucleotides and polypeptides, together or sequentially (for example, priming with a polynucleotide composition and boosting with a polypeptide composition). The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In certain embodiments, administration of the polynucleotides and polypeptides described herein is used to activate T cells. In addition to the practical advantages of simplicity of construction and modification, administration of polynucleotides encoding mutant NS polypeptides results in the synthesis of a mutant NS polypeptide in the host. Thus, these immunogens are presented to the host immune system with native post-translational modifications, structure, and conformation. The polynucleotides are preferably injected intramuscularly to a large mammal, such as a human, at a dose of 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg/kg.

The proteins and/or polynucleotides can be administered either to a mammal which is not infected with an HCV or can be administered to an HCV-infected mammal. The particular dosages of the polynucleotides or fusion proteins in a composition or will depend on many factors including, but not limited to the species,

age, and general condition of the mammal to which the composition is administered, and the mode of administration of the composition. An effective amount of the composition of the invention can be readily determined using only routine experimentation. *In vitro* and *in vivo* models can be employed to identify appropriate doses. Generally, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg will be administered to a large mammal, such as a baboon, chimpanzee, or human. If desired, co-stimulatory molecules or adjuvants can also be provided before, after, or together with the compositions.

Antibodies and Diagnostics

5

10

15

20

25

30

.

Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. Techniques for raising anti-idiotype antibodies are known in the art. See, e.g., Grzych (1985), Nature 316:74; MacNamara et al. (1984), Science 226:1325, Uytdehaag et al (1985), J. Immunol. 134:1225. These anti-idiotype antibodies may also be useful for treatment and/or diagnosis of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

An immunoassay for viral antigen may use, for example, a monoclonal antibody directed towards a viral epitope, a combination of monoclonal antibodies directed towards epitopes of one viral polypeptide, monoclonal antibodies directed towards epitopes of different viral polypeptides, polyclonal antibodies directed towards the same viral antigen, polyclonal antibodies directed towards different viral antigens or a combination of monoclonal and polyclonal antibodies.

Immunoassay protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide. The labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also

known. Examples of which are assays which utilize biotin and avidin, and enzymelabeled and mediated immunoassays, such as ELISA assays.

An enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

15

20

25

30

The HCV fusion proteins, such as NS3 mutant and core fusion proteins, can also be used to produce HCV-specific polyclonal and monoclonal antibodies. HCV-specific polyclonal and monoclonal antibodies specifically bind to HCV antigens.

Polyclonal antibodies can be produced by administering the fusion protein to a mammal, such as a mouse, a rabbit, a goat, or a horse. Serum from the immunized animal is collected and the antibodies are purified from the plasma by, for example, precipitation with ammonium sulfate, followed by chromatography, preferably affinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art.

Monoclonal antibodies directed against HCV-specific epitopes present in the fusion proteins can also be readily produced. Normal B cells from a mammal, such as a mouse, immunized with, e.g., a mutant NS3 polypeptide or NS-core fusion protein can be fused with, for example, HAT-sensitive mouse myeloma cells to produce hybridomas. Hybridomas producing HCV-specific antibodies can be identified using

RIA or ELISA and isolated by cloning in semi-solid agar or by limiting dilution.

Clones producing HCV-specific antibodies are isolated by another round of screening.

Antibodies, either monoclonal and polyclonal, which are directed against HCV epitopes, are particularly useful for detecting the presence of HCV or HCV antigens in a sample, such as a serum sample from an HCV-infected human. An immunoassay for an HCV antigen may utilize one antibody or several antibodies. An immunoassay for an HCV antigen may use, for example, a monoclonal antibody directed towards an HCV epitope, a combination of monoclonal antibodies directed towards epitopes of one HCV polypeptide, monoclonal antibodies directed towards epitopes of different HCV polypeptides, polyclonal antibodies directed towards the same HCV antigen, polyclonal antibodies directed towards different HCV antigens, or a combination of monoclonal and polyclonal antibodies. Immunoassay protocols may be based, for example, upon competition, direct reaction, or sandwich type assays using, for example, labeled antibody. The labels may be, for example, fluorescent, chemiluminescent, or radioactive.

The polyclonal or monoclonal antibodies may further be used to isolate HCV particles or antigens by immunoaffinity columns. The antibodies can be affixed to a solid support by, for example, adsorption or by covalent linkage so that the antibodies retain their immunoselective activity. Optionally, spacer groups may be included so that the antigen binding site of the antibody remains accessible. The immobilized antibodies can then be used to bind HCV particles or antigens from a biological sample, such as blood or plasma. The bound HCV particles or antigens are recovered from the column matrix by, for example, a change in pH.

Methods of Eliciting Immune Responses

5

10

15

20

25

30

HCV-specific T cells that are activated by the above-described polypeptides, expressed *in vivo* or *in vitro* preferably recognize an epitope of an HCV polypeptide such as a mutant NS3 polypeptide, including an epitope of a mutant HCV polypeptide. HCV-specific T cells can be CD8⁺ or CD4⁺.

HCV-specific CD8⁺ T cells preferably are cytotoxic T lymphocytes (CTL) which can kill HCV-infected cells that display NS3, NS4, NS5a, NS5b epitopes complexed with an MHC class I molecule. HCV-specific CD8⁺ T cells may also

express interferon-γ (IFN-γ). HCV-specific CD8⁺ T cells can be detected by, for example, ⁵¹Cr release assays. ⁵¹Cr release assays measure the ability of HCV-specific CD8⁺ T cells to lyse target cells displaying an nonstructural (e.g., mutant NS) epitope. HCV-specific CD8⁺ T cells which express IFN-γ can also be detected by immunological methods, preferably by intracellular staining for IFN-γ after *in vitro* stimulation with a mutant NS polypeptide.

HCV-specific CD4⁺ cells activated by the above-described polypeptides, expressed *in vivo* or *in vitro*, and combinations of the individual components of these proteins, preferably recognize an epitope of a mutant non-structural polypeptide, including an epitope of a mutant protein, that is bound to an MHC class II molecule on an HCV-infected cell and proliferate in response to stimulating mutant peptides.

HCV-specific CD4⁺ T cells can be detected by a lymphoproliferation assay. Lymphoproliferation assays measure the ability of HCV-specific CD4⁺ T cells to proliferate in response to an epitope.

Mutant NS (or fusions thereof with core, envelope or other viral polypeptides) can be used to activate HCV-specific T cells either *in vitro* or *in vivo*. Activation of HCV-specific T cells can be used, *inter alia*, to provide model systems to optimize CTL responses to HCV and to provide prophylactic or therapeutic treatment against HCV infection. For *in vitro* activation, proteins are preferably supplied to T cells via a plasmid or a viral vector, such as an adenovirus vector, as described above.

Polyclonal populations of T cells can be derived from the blood, and preferably from peripheral lymphoid organs, such as lymph nodes, spleen, or thymus, of mammals that have been infected with an HCV. Preferred mammals include mice, chimpanzees, baboons, and humans. The HCV serves to expand the number of activated HCV-specific T cells in the mammal. The HCV-specific T cells derived from the mammal can then be restimulated *in vitro* by adding HCV epitopic peptides to the T cells. The HCV-specific T cells can then be tested for, *inter alia*, proliferation (*e.g.*, lymphoproliferation assays known in the art), the production of IFN-γ, and the ability to lyse target cells displaying HCV NS epitopes *in vitro*.

5

10

15

20

25

The following examples are meant to illustrate the invention and are not meant to limit it in any way. Those of ordinary skill in the art will recognize modifications within the spirit and scope of the invention as set forth herein.

5 EXAMPLES

Example 1: Constructs

10

15

20

25

30

<u>pCMV-II</u>: pCMV-II (Figure 7, SEQ ID NO:5) was created to contain the human CMV promoter, enhancer, intron A, polylinker and the bovine growth hormone terminator in a deleted-pUC backbone (Life Technologies).

pT7-HCV: pT7-HCV was created in a polylinker-modified pUC vector to contain full-length HCV cDNA preceded by a synthetic T7 promoter. pT7-HCV also contains the complete 5' UTR and the poly A version of the 3' UTR.

pCMV.ΔNS35: To generate pCMV.ΔNS35 (Figure 5, SEQ ID NO:3), a two step procedure was undertaken. First, a PCR product was generated from pT7-HCV that corresponded to the following: a 5' EcoRI site, followed by the Kozak sequence of ACCATGG; the initiator ATG followed by amino acid #1242 and continuing to the StuI site. Second, the StuI to XbaI fragment from a full-length genomic clone was isolated. The genomic clone consisted of the T7 promoter fused to the full-length HCV cDNA with the poly A version of the 3' end, in a pUC vector. Finally, the EcoRI-StuI and StuI-XbaI fragments were ligated into the pCMV-II expression vector, transformed into HB101 competent cells and plated onto ampicillin (100 μg/ml). Miniprep analyses led to the identification of the desired clone which was amplified on a larger scale using a Quigen Gigaprep kit following the manufacturer's specifications. The resulting clone was named pCMV.ΔNS35 (Figure 5, SEQ ID NO:3).

pd.ΔNS3NS5: As shown schematically in Figure 10, the yeast expression plasmid pd.ΔNS3NS5 (SEQ ID NO:8) was constructed using restriction fragments obtained from the mammalian expression plasmid pCMV.KM.ΔNS35.

pCMV.KM.ΔNS35 is identical to pCMV.ΔNS35 (Figure 5, SEQ ID NO:3) except that it contains a kanamycin resistance gene in the viral backbone. pCMV.KM.ΔNS35 was digested with EcoRI and NheI to obtain 2895bp EcoRI-NheI fragment. EcoRI-NheI

fragment was ligated into pRSET HindIII-NheI subcloning vector with oligos (HE) from HindIII to EcoRI. After sequence verification, pRSETHindIII-NheI #6 was digested with HindIII and NheI to obtain a 2908bp HindIII-NheI fragment.

5

10

15

20

25

30

pCMV.KM. \(\Delta\)NS35 was linearized with XbaI and ligated with synthetic oligos (XS) from XbaI-SalI. The ligation was digested with NheI and SalI to obtain 2481bp NheI-SalI fragment. The fragment was ligated into pET3a NheI-SalI subcloning vector. After sequence verification, pET3a NheI-SalI #2 was digested with NheI and SalI to obtain a 2481bp NheI-SalI fragment. BamHI-HindIII ADH2/GAPDH promoter fragment was then ligated with HindIII-NheI and NheI-SalI fragments into pBS24.1 BamHI-SalI yeast expression vector.

pd.ΔNS3NS5.PJ: pd.ΔNS3NS5.PJ (Figures 13 and 14; SEQ ID NO:10) was generated to create a "perfect junction" at the 5' and 3' end of the HCV coding region. At the 5' end of pd.ΔNS3NS5, there were 6 extra bases between the yeast ADH2/GAPDH promoter and the ATG of the polypeptide. At the 3' end, there were 52 bases of untranslated sequence between the stop codon of the polypeptide and the α-factor terminator in the yeast expression vector. pd.ΔNS3NS5.PJ was created by digesting pd.ΔNS3NS5 #17 with ScaI and SphI to obtain 4963bp ScaI-SphI fragment. pd.NS5b3011 was digested with SphI and SalI to obtain a 321bp SphI-SalI fragment which gave the "perfect junction" at the 3' end of the polypeptide. The ScaI-SphI and SphI-SalI fragments were ligated into pSP72 HindIII-SalI subcloning vector with synthetic oligos from HindIII-ScaI(HS) for the "perfect junction" at the 5' end.

The region of synthetic sequence in pSP72 HindIII-SalI clone# 6 was verified. pSP72 HindIII-SalI clone#6 was digested with HindIII and BlnI or with BlnI and SalI to obtain 2441bp HindIII-BlnI and 2895bp BlnI-SalI fragments, respectively. The BamHI-HindIII ADH2/GAPDH promoter fragment was ligated to HindIII-BlnI and BlnI-SalI fragments into pBS24.1 BamHI-SalI yeast expression vector.

pd.ΔNS3NS5.PJ.core121RT and pd.Δ NS3NS5.PJ.core173RT were generated and encode HCV core aa 1-121 at the C-terminus of the ΔNS3NS5 polypeptide (designated pd.ΔNS3NS5.PJ.core121RT, SEQ ID NO:12) and core aa 1-173 at the C-terminus of the ΔNS3NS5 polypeptide (designated pd.ΔNS3NS5.PJ.core173RT, SEQ ID NO:14). The core sequence had aa 9 mutated from Lys to Arg and aa 11 mutated

from Asn to Thr, designated as core 121RT or 173RT.

15

20

pd.ΔNS3NS5.PJ.core121RT and pd.ΔNS3NS5.PJ.core173RT: To generate pd.ΔNS3NS5.PJ.core121RT (Figure 17, SEQ ID NO:12) and pd.ΔNS3NS5.PJ.core173RT (Figure 18, SEQ ID NO:14). As shown in Figure 16, a
 NotI-Sal HCVcore121RT and HCVcore173RT were amplified by PCR, from an E. coli expression plasmid, pSODCF2.HCVcore191RT #2. Either the core 121RT Not-SalI PCR product or the core 173RT Not-SalI PCR product were ligated into a pT7Blue2 PstI-SalI subcloning vector with synthetic oligos (PN) from PstI to NotI. After sequence confirmation, pT7Blue2core121RT clone#9 and pT7Blue2core173RT
 clone#11 was digested with PstI and SalI to obtain 403bp and 559bp PstI-SalI fragments, respectively, for further cloning.

A 121bp NotI-PstI fragment from pSP72 HindIII-SalI clone #6 was isolated as described above during the cloning of pd. \(\Delta NS3NS5.PJ. \) NotI-PstI and PstI-SalI fragments were assembled into a vector made by digesting pd. NS3NS5.PJ clone#5 (described above) with NotI and SalI.

ΔNS3NS5 and Core 140 and Core 150: An HCV core epitope was found which elicits CTLs in baboons (HCV core aa 121-135). Since pd.ΔNS3NS5.PJ.core121RT ends right before this potentially important epitope and was expressed better than the longer pd.ΔNS3NS5.PJ.core173RT construct (Example 2), two intermediate constructs were made which include this epitope, possibly giving intermediate expression levels. The two new constructs fused HCV core aa 1-140 or HCV core aa1-150 to the C terminus of ΔNS3NS5.PJ.

pd.ΔNS3NS5.PJ.core140RT (Figure 21, SEQ ID NO:16) and
pd.ΔNS3NS5.PJ.core150RT (Figure 22, SEQ ID NO:18): As shown in Figure 20, a
25 PstI-SalI HCVcore140RT and a PstI-SalIHCVcore150RT fragment were amplified by
PCR from pd.ΔNS3NS5.PJ.core173RT clone #16. Ligate either HCV core PstI-SalI
PCR products into pT7Blue2 PstI-SalI subcloning vector. After sequence
confirmation, pT7Blue2core140RT clone#22 and pT7Blue2core150RT clone#26 were
digested with PstI-SalI to obtain 460bp and 490bp PstI-SalI fragments, respectively, for
30 further cloning.

A 121bp NotI-PstI fragment was isolated from pSP72 HindIII-SalI clone #6 (as described above during the cloning of pd.ΔNS3NS5.PJ. NotI-PstI and PstI-SalI fragments were assembled into a vector made by digesting pd.ΔNS3NS5.PJ clone#5 (described above) with NotI and SalI.

5

10

15

20

30

Example 2: Protein Expression

Various of the constructs described herein, encoding HCV-1 ΔNS3 to NS5 antigen (aa 1242-3011), were expressed in yeast. *S. cerevisiae* strain AD3 was transformed with pd.ΔNS3NS5 and checked for expression. A stained protein band at the expected molecular weight of 194 kD was not observed (Figure 12). Strain AD3 was also transformed with pd.ΔNS3NS5.PJ clone #5 and checked for expression. A protein band of the expected molecular weight of 194kD was detected (Figure 15).

Strain AD3 was transformed with pd. \(\Delta NS3NS5.PJ. \) core121RT clone #6 and pd. \(\Delta NS3NS5.PJ. \) core173RT clone #15 and checked for expression. Protein bands of the expected molecular weight of 206kD and 210kD, respectively, were observed. Expression levels of the pd. \(\Delta NS3NS5.PJ. \) core173RT construct were much less than that of the pd. \(\Delta NS3NS5.PJ. \) core121RT construct. (See Figure 19). Thus, there is a correlation of protein expression levels and the length of HCV core.

Strain AD3 were transformed with pd.ΔNS3NS5.PJ.core140RT clone# 29 and pd.ΔNS3NS5.PJ.core150RT clone#35 and checked for expression. Bands of the expected molecular weights of 208kD and 209kD were seen by stain at levels close to those of pd.ΔNS3NS5core173RT (Figure 23).

Example 3: Eliciting Immune Responses

25 A. Immunization

To evaluate the immunogenicity of the mutant NS polypeptides, studies using guinea pigs, rabbits, mice, rhesus macaques and/or baboons are performed. The studies are structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed by protein immunization by PLG particles. Immunization is intramuscular or mucosally.

B. Humoral Immune Response

5

10

15

20

25

30

The humoral immune response is checked in serum specimens from immunized animals with anti-NS antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. Briefly, serum from immunized animals is screened for antibodies directed against the NS or mutant NS proteins. Wells of ELISA microtiter plates are coated overnight with the selected HCV protein and washed four times; subsequently, blocking is done with PBS-0.2% Tween (Sigma). After removal of the blocking solution, diluted mouse serum is added. Sera are tested at various dilutions. Microtiter plates are washed and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates are washed and 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) is added per well. The optical density of each well is measured. Titers are typically reported as the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.). Similarly, generation of neutralization of binding (NOB) antibodies can be measured by methods known in the art.

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) is evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Briefly, spleen cells (Effector cells, E) are obtained from the BALB/c mice immunized, cultured, restimulated, and assayed for CTL activity against HCV peptide-pulsed target cells. Cytotoxic activity is measured in a standard ⁵¹Cr release assay.

Example 4: Immunization with PLG-delivered DNA.

The polylactide-co-glycolide (PLG) polymers are obtained from Boehringer Ingelheim, U.S.A. The PLG polymer is RG505, which has a copolymer ratio of 50/50 and a molecular weight of 65 kDa (manufacturers data). Cationic microparticles with adsorbed DNA are prepared using a modified solvent evaporation process, essentially as described in Singh et al., *Proc. Natl. Acad. Sci. USA* (2000) 97:811-816. Briefly, the microparticles are prepared by emulsifying a 5% w/v polymer solution in methylene chloride with PBS at high speed using an IKA homogenizer. The primary emulsion is

then added to distilled water containing cetyl trimethyl ammonium bromide (CTAB) (0.5% w/v). This results in the formation of a w/o/w emulsion which was stirred at room temperature, allowing the methylene chloride to evaporate. The resulting microparticles are washed in distilled water by centrifugation and freeze dried.

Following preparation, washing and collection, DNA is adsorbed onto the microparticles by incubating cationic microparticles in a solution of DNA. The microparticles are then separated by centrifugation, the pellet washed with TE buffer and the microparticles are freeze dried, resuspended and administered to animals. Antibody titers are measured by ELISA assays.

10

22 -

What is claimed is:

20

 An isolated mutant non-structural ("NS") HCV polypeptide comprising a polypeptide having a mutation in the catalytic domain of NS3, wherein said mutation
 functionally disrupts the catalytic domain.

- 2. The polypeptide of claim 1, wherein the mutation comprises a deletion.
- 3. The polypeptide of claim 1, wherein the mutation comprises a substitution.
 - 4. The polypeptide of any of claims 1-3, wherein said NS polypeptide comprises NS3, NS4 and NS5.
- 15 5. The polypeptide of any of claims 1-3, wherein said NS polypeptide consists of NS3, NS4 and NS5.
 - 6. The polypeptide of any of claims 1-3, wherein said NS polypeptide consists of NS3 and NS5.
 - 7. The polypeptide of claim 6, wherein NS5 consists of NS5a.
 - 8. The polypeptide of claim 6, wherein NS5 consists of NS5b.
- 25 9. The polypeptide of any of claims 1-3, wherein said NS polypeptide consists of NS3 and NS4.
 - 10. The polypeptide of claim 9, wherein NS4 consists of NS4a.
- The polypeptide of claim 9, wherein NS4 consists of NS4b.

12. The polypeptide of claim 4, further comprising a second viral polypeptide that is not NS3, NS4, or NS5 of HCV.

- 13. The polypeptide of claim 12, wherein the second viral polypeptide comprises an HCV Core polypeptide ("C"), or fragment thereof.
 - 14. The polypeptide of claim 13, wherein the C polypeptide is truncated.
- 15. The polypeptide of claim 14, wherein the truncation is at amino acid 10 121.
 - 16. The polypeptide of claim 12, wherein the polypeptide further comprises an HCV envelope protein ("E").
- 15 17. The polypeptide of claim 16, wherein the E is E1.
 - 18. The polypeptide of claim 16, wherein the E is E2.
 - 19. A composition comprising
 - (a) the polypeptide of any one of claims 1-18; and
 - (b) a pharmaceutically acceptable excipient.
 - 20. An isolated and purified polynucleotide which encodes the mutant HCV polypeptide according to any one of claims 1-18.

25

20

- 21. A composition comprising
- (a) the isolated purified polynucleotide of claim 20; and
- (b) a pharmaceutically acceptable excipient.
- The composition of claim 21, wherein the polynucleotide is DNA.

The composition of claim 21, wherein the polynucleotide is in a

23.

		plasmid.
5	24.	An expression vector comprising the polynucleotide of claim 20.
J	25.	An expression vector comprising the polynucleotide of SEQ ID NO:8.
	26.	A host cell comprising the polynucleotide of claim 20.
10	27.	The host cell of claim 26, wherein the cell is a yeast cell.
	28.	The host cell of claim 26, wherein the cell is a mammalian cell.
15	29.	The host cell of claim 26, wherein the cell is an insect cell.
13	30.	The host cell of claim 26, wherein the cell is a plant cell.
	31. sequence of S.	The host cell of claim 26, wherein the polynucleotide comprises the
20	sequence of B	LQ ID 110.0.
20	32.	The polypeptide of claim 1, wherein the polypeptide further comprises
	SEQ ID NO:9	.
•	33.	A method of preparing a mutant NS HCV polypeptide, wherein the
25	method comp	rises the steps of:
		a. transforming a host cell with an expression vector according to claim 24, under conditions wherein the polypeptide is expressed and
30		b. isolating the polypeptide.

- 34. The method of claim 33, wherein the host cell is a yeast cell.
- 35. The method of claim 33, wherein the host cell is a mammalian cell.
- 5 36. The method of claim 33, wherein the host cell is an insect cell.
 - 37. The method of claim 33, wherein the host cell is a plant cell.
- 38. An antibody that specifically binds to a polypeptide of any of claims 110 18.
 - 39. The antibody of claim 38, wherein the antibody is a monoclonal antibody.
- 15 40. The antibody of claim 38, wherein the antibody is a purified polyclonal antibody.
 - 41. A method of eliciting an immune response in a subject, comprising the step of administering to the subject a polypeptide of any of claims 1-18.

20

4-.::

42. A method of eliciting an immune response in a subject, comprising the step of administering to the subject a polynucleotide of claim 20.

FIGURE 1

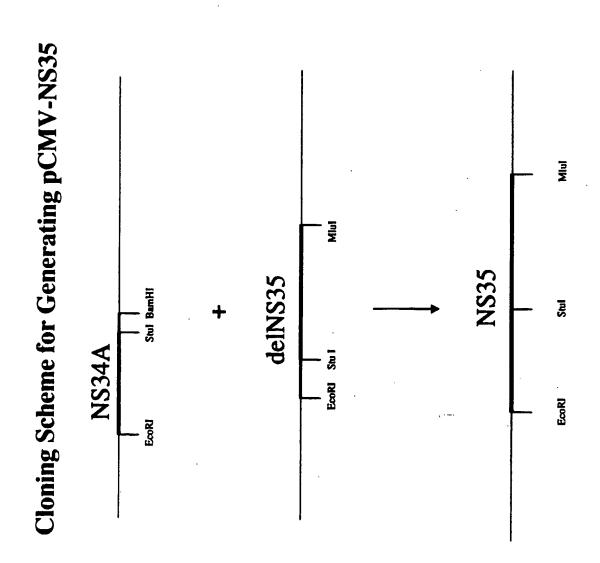
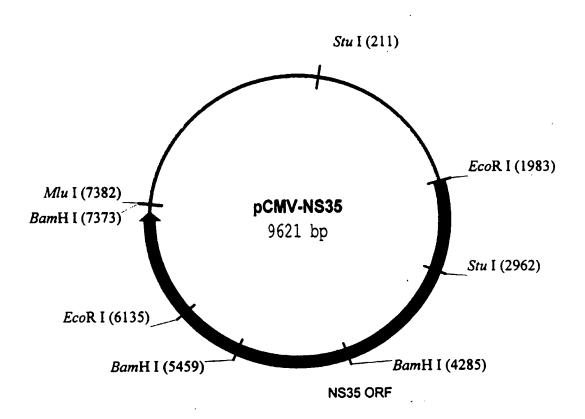


FIGURE 2



pCMV-NS35 2/100

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT
	AGCGCGCAAA	GCCACTACTG	CCACTTTTGG	AGACTGTGTA	CGTCGAGGGC	CTCTGCCAGT	GTCGAACAGA	CATTCGCCTA
81	GCCGGGAGCA	GACAAGCCCG	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA
	CGGCCCTCGT	CTGTTCGGGC	AGTCCCGCGC	AGTCGCCCAC	AACCGCCCAC	AGCCCCGACC	GAATTGATAC	GCCGTAGTCT
					Stu			
161	GCAGATTGTA	CTGAGAGTGC	ACCATATGAA	GCTTTTTGCA	AAAGCCTAGG	CCTCCAAAAA	AGCCTCCTCA	CTACTTCTGG
	CGTCTAACAT	GACTCTCACG	TGGTATACTT	CGAAAAACGT	TTTCGGATCC	GGAGGTTTTT	TCGGAGGAGT	GATGAAGACC
241	AATAGCTCAG	AGGCCGAGGC	GGCCTCGGCC	TCTGCATAAA	TAAAAAAAT	TAGTCAGCCA	TGGGGCGGAG	AATGGGCGGA
	TTATCGAGTC	TCCGGCTCCG	CCGGAGCCGG	AGACGTATTT	ATTTTTTTA	ATÇAGTCGGT	ACCCCGCCTC	TTACCCGCCT
321	ACTGGGCGGG	GAGGGAATTA	TTGGCTATTG	GCCATTGCAT	ACGTTGTATC	TATATCATAA	TATGTACATT	TATATTGGCT
	TGACCCGCCC	CTCCCTTAAT	AACCGATAAC	CGGTAACGTA	TGCAACATAG	ATATAGTATT	ATACATGTAA	ATATAACCGA
401	CATGTCCAAT	ATGACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT
	GTACAGGTTA	TACTGGCGGT	ACAACTGTAA	CTAATAACTG	ATCAATAATT	ATCATTAGTT	AATGCCCCAG	TAATCAAGTA
481	AGCCCATATA	TGGAGTTCCG	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	CCCGCCCATT
	TCGGGTATAT	ACCTCAAGGC	GCAATGTATT	GAATGCCATT	TACCGGGCGG	ACCGACTGGC	GGGTTGCTGG	GGGCGGGTAA
561	GACGICAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT
	CTGCAGTTAT	TACTGCATAC	AAGGGTATCA	TTGCGGTTAT	CCCTGAAAGG	TAACTGCAGT	TACCCACCTC	ATAAATGCCA
641	AAACTGCCCA	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTCCGCCC	CCTATTGACG	TCAATGACGG	TAAATGGCCC
	TTTGACGGGT	GAACCGTCAT	GTAGTTCACA	TAGTATACGG	TTCAGGCGGG	GGATAACTGC	AGTTACTGCC	ATTTACCGGG
721	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	CGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC
	CGGACCGTAA	TACGGGTCAT	GTACTGGAAT	GCCCTGAAAG	GATGAACCGT	CATGTAGATG	CATAATCAGT	AGCGATAATG
801	CATGGTGATG GTACCACTAC	CGGTTTTGGC GCCAAAACCG	AGTACACCAA TCATGTGGTT	TGGGCGTGGA ACCCGCACCT	TAGCGGTTTG	ACTCACGGGG TGAGTGCCCC	ATTTCCAAGT TAAAGGTTCA	CTCCACCCA GAGGTGGGGT
881	TTGACGTCAA AACTGCAGTT	TGGGAGTTTG ACCCTCAAAC	TTTTGGCACC	AAAATCAACG TTTTAGTTGC	GGACTTTCCA CCTGAAAGGT	AAATGTCGTA TTTACAGCAT	ATAACCCCGC TATTGGGGCG	CCCGTTGACG GGGCAACTGC
961	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATA	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG
	GTTTACCCGC	CATCCGCACA	TGCCACCCTC	CAGATATAT	CGTCTCGAGC	AAATCACTTG	GCAGTCTAGC	GGACCTCTGC
1041	CCATCCACGC	TGTTTTGACC	TCCATAGAAG AGGTATCTTC	ACACCGGGAC TGTGGCCCTC	CGATCCAGCC GCTAGGTCGG	TCCGCGGCCG AGGCGCCGGC	GGAACGGTGC	ATTGGAACGC TAACCTTGCG
1121	GGATTCCCCG CCTAAGGGGC	TGCCAAGAGT ACGGTTCTCA	GACGTAAGTA CTGCATTCAT	CCGCCTATAC	ACTCTATAGO TGAGATATCO	CACACCCCTT GTGTGGGGA	TGGCTCTTAT A ACCGAGAATA	GCATGCTATA CGTACGATAT
1201	CTGTTTTTGG GACAAAAACG	CTTGGGGCCT	ATACACCCCC TATGTGGGGG	GCTCCTTATO CGAGGAATAO	GATATCCACT	TGGTATAGCT ACCATATCG	TAGCCTATAG	GTGTGGGTTA CACACCCAAT
1281	TTGACCATTA AACTGGTAAT	TTGACCACTO	CCCTATTGGT GGGATAACC	GACGATACT	TCCATTACTA A AGGTAATGAT	ATCCATAACA TAGGTATTG	A TGGCTCTTTC T ACCGAGAAAC	GCACAACTAT GGTGTTGATA
1361	CTCTATTGGC GAGATAACCC	TATATGCCAF	A TACTCTGTCC	TTCAGAGAC AAGTCTCTG	I GACACGGACI A CTGTGCCTGA	CTGTATTTT	A TGTCCTACCO	G GTCCATTTAT C CAGGTAAATA

pCMV-NS35

TATTTACAAA ATAAATGTTT	TTCACATATA AAGTGTATAT	CAACAACGCC GTTGTTGCGG	GTCCCCCGTG CAGGGGGCAC	CCCGCAGTTT GGGCGTCAAA	TTATTAAACA AATAATTTGT	TAGCGTGGGA ATCGCACCCT	TCTCCGACAT AGAGGCTGTA
CTCGGGTACG GAGCCCATGC	TGTTCCGGAC ACAAGGCCTG	ATGGGCTCTT TACCCGAGAA	CTCCGGTAGC GAGGCCATCG	GGCGGAGCTT CCGCCTCGAA	CCACATCCGA GGTGTAGGCT	GCCCTGGTCC CGGGACCAGG	CATCCGTCCA GTAGGCAGGT
GCGGCTCATG CGCCGAGTAC	GTCGCTCGGC CAGCGAGCCG	AGCTCCTTGC TCGAGGAACG	TCCTAACAGT AGGATTGTCA	GGAGGCCAGA CCTCCGGTCT	CTTAGGCACA GAATCCGTGT	GCACAATGCC CGTGTTACGG	CACCACCACC GTGGTGGTGG
AGTGTGCCGC TCACACGGCG	ACAAGGCCGT TGTTCCGGCA	GGCGGTAGGG CCGCCATCCC	TATGTGTCTG ATACACAGAC	AAAATGAGCT TTTTACTCGA	CGGAGATTGG GCCTCTAACC	GCTCGCACCT CGAGCGTGGA	GGACGCAGAT CCTGCGTCTA
GGAAGACTTA CCTTCTGAAT	AGGCAGCGGC TCCGTCGCCG	AGAAGAAGAT TCTTCTTCTA	GCAGGCAGCT CGTCCGTCGA	GAGTTGTTGT CTCAACAACA	ATTCTGATAA TAAGACTATT	GAGTCAGAGG CTCAGTCTCC	TAACTCCCGT ATTGAGGGCA
TGCGGTGCTG ACGCCACGAC	TTAACGGTGG AATTGCCACC	AGGGCAGTGT TCCCGTCACA	AĞTCTGAGCA TCAGACTCGT	GTACTCGTTG CATGAGCAAC	CTGCCGCGCG GACGGCGCGC	CGCCACCAGA GCGGTGGTCT	CATAATAGCT GTATTATCGA
· ·						ECORI	M A A
GACAGACTAA CTGTCTGATT	CAGACTGTTC GTCTGACAAG	CTTTCCATGG GAAAGGTACC	GTCTTTTCTG CAGAAAAGAC	CAGTCACCGT GTCAGTGGCA	CGTCGACCTA GCAGCTGGAT	AGAATTCACC TCTTAAGTGG	ATGGCTGCAT TACCGACGTA
ATGCAGCTCA	GGGCTATAAG	GTGCTAGTAC	TCAACCCCTC	TGTTGCTGCA	ACACTGGGCT	TTGGTGCTTA	M S K CATGTCCAAG GTACAGGTTC
GCTCATGGGA	TCGATCCTAA	I R T CATCAGGACC GTAGTCCTGG	GGGGTGAGAA	CAATTACCAC	G S P TGGCAGCCCC ACCGTCGGGG	ATCACGTACT	CCACCTACGG
K F L CAAGTTCCTT GTTCAAGGAA	GCCGACGGCG	GGTGCTCGGG	GGGCGCTTAT	GACATAATAA	TTTGTGACGA	C H S GTGCCACTCC CACGGTGAGG	T D A ACGGATGCCA TGCCTACGGT
T S I L CATCCATCTT GTAGGTAGAA	G I G GGGCATTGGC CCCGTAACCG	ACTGTCCTTG	ACCAAGCAGA	GACTGCGGGG	GCGAGACTGG	TTGTGCTCGC	T A T CACCGCCACC GTGGCGGTGG
CCTCCGGGCT	CCGTCACTGT	P H P GCCCCATCCC CGGGGTAGGG	AACATCGAGG	AGGTTGCTCT	S T T GTCCACCACC CAGGTGGTGG	GGAGAGATCC	CTTTTTACGG
CAAGGCTATC	CCCCTCGAAG	TAATCAAGGG	GGGGAGACAT	CTCATCTTCT	GTCATTCAAA	K K C GAAGAAGTGC CTTCTTCACG	D E L GACGAACICG CTGCTTGAGC
		GGCATCAATG	CCGTGGCCTA	CTACCGCGGT	CTTGACGTGT	CCGTCATCCC	
GATGTTGTCG	TCGTGGCAAC	D A L CGATGCCCTC GCTACGGGAG	ATGACCGGCT	ATACCGGCGA	F D S CTTCGACTCG GAAGCTGAGC	GTGATAGACT.	GCAATACGTG
	CTCGGGTACG GAGCCCATGC GCGCCCATGC GCGCCGAGTAC AGTGTGCCGC GGAAGACTTA CCTTCTGAAT TGCGGTGCTG ACGCCACGAC GACAGACTAA CTGTCTGATT Y A A Q ATGCAGCTCA TACGTCGAGT A H G I GCTCATGGGA CGAGTACCCT K F L CAAGTTCCTT GTTCAAGGAA T S I L CATCCATCTT GTAGGTAGAA P P G S CCTCCGGGCT GGAGGCCCGA K A I CAAGGCTATC GTTCCGATAG A A K L CCGCAAAGCT GGCGTTTCGA D V V V GATGTTGTCG	CTCGGGTACG TGTTCCGGAC GAGCCCATGC ACAAGGCCTG GCGGCTCATG GTCGCTCGGC CGCCGAGTAC CAGCGAGCCG AGTGTGCCGC ACAAGGCCGT TCACACGGCG TGTTCCGGCA GGAAGACTTA AGGCAGCGGC CCTTCTGAAT TCCGTCGCCG TGCGGTGCTG TTAACGGTGG ACGCCACGAC AATTGCCACC GACAGACTAA CAGACTGTTC CTGTCTGATT GTCTGACAAG Y A A Q G Y K ATGCAGCTCA GGGCTATAAG TACGTCGAGT CCCGATATTC A H G I D P N GCTCATGGGA TCGATCCTAA CGAGTACCT AGCTAGGATT K F L A D G C CAAGTTCCTT GCCGACGGCG GTTCAAGGAA CGGCTGCCGC T S I L G I G CATCCATCTT GGCGATTGGC GTAGGTAGAA CCCGTAACCG P P G S V T V CCTCCGGGCT CCGTCACTGT GGAGGCCCGA GGCAGTGACA K A I P L E V CAAGGCTATC CCCCTCGAAG GTTCCGATAG GGGGGAGCTTC A A K L V A L CCGCAAAGCT GGTCGCATTG GGCGTTTCGA CCAGCGTTAC D V V V V A T GATGTTGTCG TCGTGGCAAC	CTCGGGTACG TGTTCCGGAC ATGGGCTCTT GAGCCCATGC ACAAGGCCTG TACCCGAGAA GCGGCTCATG GTCGCTCGGC AGCTCCTTGC CGCCGAGTAC CAGCGAGCCG TCGAGGAACG AGTGTGCCGC ACAAGGCCGT GGCGGTAGGG TCACACGGCG TGTTCCGGCA CCGCCATCCC GGAAGACTTA AGGCAGCGGC AGAAGAAGAT CCTTCTGAAT TCCGTCGCCG TCTTCTTCTA TGCGGTGCTG TTAACGGTGG AGGGCAGTGT ACGCCACGAC AATTGCCACC TCCCGTCACA GACAGACTAA CAGACTGTTC CTTTCCATGG CTGTCTGATT GTCTGACAAG GAAAGGTACC Y A A Q G Y K V L V I ATGCAGCTCA GGGCTATAAG GTGCTAGTAC TACGTCGAGT CCCGATATTC CACGATCATG A H G I D P N I R T GCTCATGGGA TCGATCCTAA CATCAGGACC CGAGTACCCT AGCTAGGATT GTAGTCCTGG K F L A D G G C S G CAAGATTCCTT GCCGACGGCG GGTGCTCGGG GTTCAAGGAA CGGCTGCCGC CCACGAGCCC T S I L G I G T V L I CATCCATCTT GGCGACTGGC ACTGTCCTTG GTAGGTAGAA CCCGTAACCG TGACAGGAAC P P G S V T V P H P CCTCCGGGCT CCGTCACTGT GCCCCATCCC GGAGGCCCGA GGCAGTGACA CGGGGTAGGG K A I P L E V I K G CAAGGCTATC CCCCTCGAAG TAATCAAGGG GTTCCGATAG GGCGATGACA CGGGGTAGGG K A I P L E V I K G CAAGGCTATC CCCCTCGAAG TAATCAAGGG GTTCCGATAG GGCGATGACA CGGGGTAGGG K A I P L E V I K G CAAGGCTATC CCCCTCGAAG TAATCAAGGG GTTCCGATAG GGCGATCATG GGCGTTTCGA CCGCGATGCCCTC D V V V V A T D A L GATGTTGTCG TCGTGGCAAC CGGTGCCCTC	CTCGGGTACG TGTTCCGGAC ATGGGCTCTT CTCCGGTAGC GAGCCCATGC ACAAGGCCTG TACCCGAGAA GAGGCCATCG GCGGGTACC GCCCGAGTAC CAGCGAGCCG TCGAGGAACG AGGATTGTCA CAGCGAGTAC CAGCGAGCCG TCGAGGAACG AGGATTGTCA AGTGTGCCGC ACAAGGCCGT GGCGGTAGGG TATGTGTCTG TCACACGGC TGTTCCGGCA CCGCCATCCC ATACACAGAC CCTTCTGAAT TCCGTCGCCG ACAAGGACGT AGGCAGCACCT CCTTCTGAAT TCCGTCGCCG TCTTCTTCTA CGTCCGTCGA ACGCCACCA TCACACAGAC TCCCTTCTGAAT TCCGTCGCCG TCTTCTTCTA CGTCCGTCGA ACGCCACCA AATTGCCACC TCCCGTCACA TCAGACTCGT TACGCCACA AATTGCCACC TCCCGTCACA TCAGACTCGT TACGTCGAGCA AATTGCCACC TCCCGTCACA TCAGACTCGT TACGTCGAGT ACGCCACAACAAG GAAAGGTACC CAGAAAAAGAC CAGACACAAG GAAAGGTACC CAGAAAAAGAC TACGCCACCA TCAGACCCCTC TACGTCGAGT CCCGGATATTC CACGATCATG AGTTGGGGAG CGAGTACCCTA TCAGACCCCTC TACGTCGAGT TCGATCCTAA CATCAGGACC GGGGTGAGAA CGAGTACCCT AGCTAGGAT TGAGTCCTGG CCCCACCTCTT CACGTCACAC TGAGTCCTTA GGTTCAAGGAC GGGGTGAGAA CGAGTACCCT AGCTAGGAT TGAGTCCTGG CCCCACCTCTT TACGTCCACAC CGAGAGACACACACACACACACACACACACACA	CTCGGGTACG TGTTCCGGC ATGGGCTCTT CTCCGGTAGC GGCGGAGCTT GAGCCCATGC ACAAGGCCT TACCCGAGAA GAGGCCATGC CCGCCTCGAA GCGCCATGC ACAAGGCCT TACCCGAGAA GAGGCCATGC CCGCCTCGAA GCGCGAGTAC CAGCGAGCCG TCGAGGAACGA GAGGCCATGC CCGCCTCGAA GCGCGAGTAC CAGCGAGCCG TCGAGGAACGA AGGATTGTCA CCTCCGGTCT AGTGTGCCGCA CAAGGCCGT GGCGGTAGGG TATGTGTCA CCTCCGGTCT CACACAGGC TGTTCCGGCA CCGCCATCCC ATACACAGAC TTTTACTCGA GGAAGACATTA AGGCAGCGGC AGAAGAAGAT GCAGGCAGCT GAGTTGTTGCCTTCTGAAT TCCGTCGCGC TCTTCTTCTA CGTCCGTCGA CTCAACAACA TCGTTGGAGCA AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAACA AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAACA AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAACA AATGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAACA TACGTCGAGTA GGCTATAAG GAAAGGTACC CAGAAAAGAAC GTCAGTGGCA AATGCCACC TCCCGTCACA TCAGACCCT TGTTGCTGACAA GAAAGGTACC CAGAAAAGAAC GTCAGTGGCA TACGTCGAGT GGCCATATAC CACGATCATGA GATTGGGGAA CAACGACGT TACGTCGAGT CCCGGATATTC CACGATCATGA GATTGGGGAA CAACGACGAT TACGTCGAGT TCAGACCCT TGTTGCTGCA TACGTCGAGT CCCGGATAACC CCGGATACCC TACGACCACC TACACCCCT TGTTACCACCCGAGAAAAGAACACA CACATCACACCAC AGAATAACAA CACCACGAGAACAA CAATTACCAC CCAGATACCCT AGCTAGGAAC TACGTCACACCAC AGAATAACAA CACCACGAGAACACA CACATCAAGAAAAA CAGCACCGAGCACC CCCGCCGCATAT CACATAATAAA GTTCAAGGAAA CAGCTCCCGC CCCCGCCCATAT CACATAATAAA GTTCAAGGAAA CAGCTCCCGC CCCCGCCGCATAT CACATAATAAA CTGTATTATAA GTTCAAGGAAA CAGCTCCCGCC CCCCGCCCATATA CTGTATTATAA GTTCAAGGAAA CAGCTCCCGCC CCCCGCCGCAATAA CTGTATTATAA GTTCAAGGAAA CAGCTCCCGCC CCCCGCCCATAA CACATCAAGAAAA CACCTCCTCT GCCCCCCCCCC	CTCGGGTACG TGTTCCGGAC ATGGGCTCTT CTCCCGTAGC GGCGGAGCTT CCACATCCGA CAGCCCATGC ACAAGGCCTG TACCCGAAAA AGAATATTTCT CTCCGGTAGC GGCGCTCGAA GGTGTAGGCT CACACCGAA GAGCCCATGC ACAAGGCCTG TACCCGAAAAAAAAAA	TCACACAGGGG TGTTCCGGCA CCGCCATCCC ATACACAGAC TTTTACTCA GCCTCTAACC CGAGGGTGGA GGAAGACTTA AGGCAGCGGC AGAAGAAGAT GCAGGCAGCT GAGTTGTTGT ATTCTGATAA GAGTCAGAGG CCTTCTGAAT TCCGTCGCCG TCTTCTTCTA CGTCCGTCGA CTCAACAACA TAAGACTATT CTCAGTCTC TGCGGTGCTG TTAACGGTGG AGGGCAGTGT AGTCTGAGCA GTACTCGTTG CTGCCGCGGC GCCCACCAGA ACGCCACGAC AATTGCCACC TCCCGTCACA TCAGACTCGT CATGAGCAAC GACGGCGCG GCGACCAGA ACGCCACAAA CAGACTGTTC CTTTCCATGG GTCTTTTCTG CAGTCACCGT CGTCGACCCTA AGAATTCACC CTGTCTGATT GTCTGACAAG GAAAGGTACC CAGAAAAGAC GTCAGTGGCA GCAGCTGGAT TCTTAAGTGG Y A A Q G Y K V L V L N P S V A A T L G F G A Y ATGCAGCCTA GGGCTATAAG GTGCTAAGTAC TCAACCCCTC TGTTGCTGCA AACCTGGGCT TTGTGGTCTTA TACGTCGAGT CCCGATATTC CACGATCATG AGTTGGGGAG ACAACGACGT TGTGACCCCGA AACCACGGAT A H G I D P N I R T G V R T I T T G S P I T Y S GCTCATGGGA TCGATCCTAA CATCAGGACC GGGGTGAGAA CAATTACCAC TGGCAGCCC ATCACGTACT CGAGTACCCT AGCTAGGAT GTAGTCCTGG GGGGCTATA GACATAATAA TTTGTGACGG TAGTGCATGA K F L A D G C S G A Y D I I C D E C H S CAAGTTCCTT GCCGACGCGC CCACCGACTCTG GGGGCTTAT GACATAATAA TTTTGTGACGG TAGTGCACTCC GTTCAAGGAA CGGCTGCGGC CCACGAGACCC CCCCGCGAATA CTGTATTATT AAACACTGTC CACGGTGAGG TTCAAGGAA CGGCTGCCGC CCACGAGACCC CCCCCGGAATA CTGTATTATTA AAACACTGCT CACGGTGAGG TTGAAGGTAAACGAA CAGGTGCCC CACCGAGACCC CCCCCGCGAATA CTGTATTATTA AAACACTGCT CACGGTGAGG GTACGGTAACCG TGACAGGAAC TGGTTCGTCT CTGACCCCCC CGCTCTGACC AACACGAGGC P P G S V T V P H P N I E E V A L S T T G E I F CCTCCGGGCT CCGTCACCG GACGACACTCC AACACCAGAG CACGTTGCTC GCCCACCC GGAAGAGTCC GGAGGCCCGA GCCACTACCG TGACAGGAACTCC TCCAACGAGA CAGGTGGGG CCTCTCTAGC GGAGGCCCGA GCCCACCTCC AACACCAGAGA CCCCCCCACCC GGAAGACTCC GGAGGCCCGA GCCCACCCCCCCCCCCCCTCTCTGA GAGGTAGGG CCCCCCCCCC

pCMV-NS35

+2 2641.	TOTORCOCAG	ACAGTCGATT	TCAGCCTTGA	CCCTACCTTC	ACCATTGAGA	I T L CAATCACGCT (GTTAGTGCGA (CCCCAAGAT G	A V S CTGTCTCCC GACAGAGGG
+2 2721	R T Q R GCACTCAACG CGTGAGTTGC	TOGGGGGCAGG	ACTGGCAGGG	GGAAGCCAGG	I Y R CATCTACAGA GTAGATGTCT	F V A P TTTGTGGCAC (AAACACCGTG (CGGGGGAGCG C	CCCTCCGGC
+2	ATGTTCGACT	S S V L CGTCCGTCCT GCAGGCAGGA	CTGTGAGTGC	TATGACGCAG	C A W GCTGTGCTTG CGACACGAAC	Y E L GTATGAGCTC CATACTCGAG	T P A E ACGCCCGCCG A TGCGGGCGGC T	GACTACAGT
+2	RLR	A Y M t	1 T P G	L P V	CQDF	LEF	W E G	V F T StuI
2881	TAGGCTACGA ATCCGATGCT	GCGTACATGA CGCATGTACT	ACACCCCGGG TGTGGGGCCC	GCTTCCCGTG CGAAGGGCAC	TGCCAGGACC ACGGTCCTGG	ATCTTGAATT TAGAACTTAA	TTGGGAGGGC C	TCTTTACAG CAGAAATGTC
+2	StuI	I D A	H F L	з отк	Q S G	E N L P	YLV	A Y Q
2961	GCCTCACTCA CGGAGTGAGT	TATAGATGCC ATATCTACGG	CACTTTCTAT GTGAAAGATA	CCCAGACAAA GGGTCTGTTT	GCAGAGTGGG CGTCTCACCC	GAGAACCTTC CTCTTGGAAG	CTTACCTGGT A	AGCGTACCAA PCGCATGGTT
+2 3041	A T V GCCACCGTGT CGGTGGCACA	CCCCTACCCC	TONACCOCCT	P P S CCCCCATCGT GGGGGTAGCA	GGGACCAGAT	W K C GTGGAAGTGT CACCTTCACA	TTGATTCGCC '	CAAGCCCAC
+2 3121		P T P CCAACACCCC GGTTGTGGGG	TCCTATACAG	L G A ACTGGGCGCT TGACCCGCGA	GTTCAGAATG	E I T L AAATCACCCT TTTAGTGGGA	GACGCACCCA (STCACCAAAT
+2 3201	Y I M T ACATCATGAC TGTAGTACTG	********	A D L GCCGACCTGG CGGCTGGACC	ACCTCCTCAC	GAGCACCTGG	V L V C GTGCTCGTTG CACGAGCAAC	GCGGCGTCCT	GGCTGCTTTG
3281	A A Y GCCGCGTATT CGGCGCATAA		* *CCCTCCCTC	V I V GTCATAGTGG CAGTATCACC	CCACGGTCGT	L S G CTTGTCCGGG GAACAGGCCC	AAGCCGGCAA	TCATACCTGA
3361			**********	CATCCAACAC	TGCTCTCAGC	H L P Y ACTTACCGTA TGAATGGCAT	CATCGAGCAA	GGGATGATGC
3441	L A E C TCGCCGAGCA AGCGGCTCG1		* **********	G L L C G GCCTCCTGCA C CGGAGGACG1	CACCGCGTCC	R Q A CGTCAGGCAGGGCAGGGCAGTCCGTC	AGGTTATCGC	P A V CCCTGCTGTC GGGACGACAG
+: 3521	Q T N CAGACCAACT GTCTGGTTGA	·			* *********	F I S CTTCATCAGT GAAGTAGTCA	GGGATACAAT	ACTTGGCGGG
3601						T A A V A CAGCTGCTGT C GTCGACGACA		
3681						A A P GCCGCCCCG GCGCGGGGGC	GTGCCGCTAC	A F V TGCCTTTGTG ACGGAAACAC

pCMV-NS35

+2 G A G L A G A A I G S V G L G K V L I D I L A G Y G A 3761 GGCGCTGGCT TAGCTGGCGC CGCCATCGGC AGTGTTGGAC TGGGGAAGGT CCTCATAGAC ATCCTTGCAG GGTATGGCGC CCGCGACCGA ATCGACCGCG GCGGTAGCCG TCACAACCTG ACCCCTTCCA GGAGTATCTG TAGGAACGTC CCATACCGCG	_
+2 G V A G A L V A F K I M S G E V P S T E D L V N L L 3841 GGGCGTGGCG GGAGCTCTTG TGGCATTCAA GATCATGAGC GGTGAGGTCC CCTCCACGGA GGACCTGGTC AATCTACTGC CCCGCACCGC CCTCGAGAAC ACCGTAAGTT CTAGTACTCG CCACTCCAGG GGAGGTGCCT CCTGGACCAG TTAGATGACG	-
+2 P A I L S P G A L V V G V V C A A I L R R H V G P G E 3921 CCGCCATCCT CTCGCCCGGA GCCCTCGTAG TCGGCGTGGT CTGTGCAGCA ATACTGCGCC GGCACGTTGG CCCGGGCGAG GGCGGTAGGA GAGCGGGCCT CGGGAGCATC AGCCGCACCA GACACGTCGT TATGACGCGG CCGTGCAACC GGGCCCGCTC	_
+2 G A V Q W M N R L I A F A S R G N H V S P T H Y V P E 4001 GGGGCAGTGC AGTGGATGAA CCGGCTGATA GCCTTCGCCT CCCGGGGGAA CCATGTTTCC CCCACGCACT ACGTGCCGGA CCCCGTCACG TCACCTACTT GGCCGACTAT CGGAAGCGGA GGGCCCCCTT GGTACAAAGG GGGTGCGTGA TGCACGGCCT	_
+2 S D A A R V T A I L S S L T V T Q L L R R L H Q W 4081 GAGCGATGCA GCTGCCGCG TCACTGCCAT ACTCAGCAGC CTCACTGTAA CCCAGCTCCT GAGGCGACTG CACCAGTGGA CTCGCTACGT CGACGGGCGC AGTGACGGTA TGAGTCGTCG GAGTGACATT GGGTCGAGGA CTCCGCTGAC GTGGTCACCT	_
+2 I S S E C T T P C S G S W L R D I W D W I C E V L S D 4161 TAAGCTCGGA GTGTACCACT CCATGCTCCG GTTCCTGGCT AAGGGACATC TGGGACTGGA TATGCGAGGT GTTGAGCGAC ATTCGAGCCT CACATGGTGA GGTACGAGGC CAAGGACCGA TTCCCTGTAG ACCCTGACCT ATACGCTCCA CAACTCGCTG	_
+2 FKTW LKA KLM PQLP GIPFVS CQRG YKG BamHI	
4241 TITAAGACCT GGCTAAAAGC TAAGCTCATG CCACAGCTGC CTGGGATCCC CTTTGTGTCC TGCCAGCGCG GGTATAAGGG AAATTCTGGA CCGATTTTCG ATTCGAGTAC GGTGTCGACG GACCCTAGGG GAAACACAGG ACGGTCGCGC CCATATTCCC	_
+2 V W R G D G I M H T R C H C G A E I T G H V K N G T 4321 GGTCTGGCGA GGGGACGGA TCATGCACAC TCGCTGCCAC TGTGGAGCTG AGATCACTGG ACATGTCAAA AACGGGACGA CCAGACCGCT CCCCTGCCGT AGTACGTGTG AGCGACGGTG ACACCTCGAC TCTAGTGACC TGTACAGTTT TTGCCCTGCT	_
+2 M R I V G P R T C R N M W S G T F P I N A Y T T G P C 4401 TGAGGATCGT CGGTCCTAGGA ACCTGCAGGA ACATGTGGAG TGGGACCTTC CCCATTAATG CCTACACCAC GGGCCCCTGT ACTCCTAGCA GCCAGGATCC TGGACGTCCT TGTACACCTC ACCCTGGAAG GGGTAATTAC GGATGTGGTG CCCGGGGACA	_
+2 T P L P A P N Y T F A L W R V S A E E Y V E I R Q V G 4481 ACCCCCTTC CTGCGCCGAA CTACACGTTC GCGCTATGGA GGGTGTCTGC AGAGGAATAC GTGGAGATAA GGCAGGTGGG TGGGGGGAAG GACGCGGCTT GATGTGCAAG CGCGATACCT CCCACAGACG TCTCCTTATG CACCTCTATT CCGTCCACCC	
+2 D F H Y V T G M T T D N L K C P C Q V P 5 P E F F T 4561 GGACTTCCAC TACGTGACGG GTATGACTAC TGACAATCTT AAATGCCCGT GCCAGGTCCC ATCGCCCGAA TTTTTCACAG CCTGAAGGTG ATGCACTGCC CATACTGATG ACTGTTAGAA TTTACGGGCA CGGTCCAGGG TAGCGGGCTT AAAAAGTGTC	
+2 E L D G V R L H R F A P P C K P L L R E E V S F R V G 4641 AATTGGACGG GGTGCGCCTA CATAGGTTTG CGCCCCCCTG CAAGCCCTTG CTGCGGGAGG AGGTATCATT CAGAGTAGGA TTAACCTGCC CCACGCGGAT GTATCCAAAC GCGGGGGGAC GTTCGGGAAC GACGCCCTCC TCCATAGTAA GTCTCATCCT	_
+2 L H E Y P V G S Q L P C E P E P D V A V L T S M L T D 4721 CTCCACGAAT ACCCGGTAGG GTCGCAATTA CCTTGCGAGC CCGAACCGGA CGTGGCCGTG TTGACGTCCA TGCTCACTGA GAGGTGCTTA TGGGCCATCC CAGCGTTAAT GGAACGCTCG GGCTTGGCCT GCACCGGCAC AACTGCAGGT ACGAGTGACT	
+2 P S H I T A E A A G R R L A R G S P P S V A S S S A 4801 TCCCTCCCAT ATAACAGCAG AGGCGGCCGG GCGAAGGTTG GCGAGGGGAT CACCCCCCTC TGTGGCCAGC TCCTCGGCTA AGGGAGGGTA TATTGTCGTC TCCGCCGGCC CGCTTCCAAC CGCTCCCCTA GTGGGGGGAG ACACCGGTCG AGGAGCCGAT	

pCMV-NS35

4881	S Q L S GCCAGCTATC CGGTCGATAG	CGCTCCATCT	L K A CTCAAGGCAA GAGTTCCGTT	CTTGCACCGC	TAACCATGAC	S P D TCCCCTGATG AGGGGACTAC	A E L I CTGAGCTCAT GACTCGAGTA	E A N AGAGGCCAAC TCTCCGGTTG
4961	L L W CTCCTATGGA GAGGATACCT	GGCAGGAGAT	GGGCGGCAAC	I T R ATCACCAGGG TAGTGGTCCC	TTGAGTCAGA	N K V AAACAAAGTG TTTGTTTCAC	V I L D GTGATTCTGG CACTAAGACC	ACTCCTTCCA
5041	TCCGCTTGTG	A E E I GCGGAGGAGG CGCCTCCTCC	ACGAGCGGGA	GATCTCCGTA	P A E CCCGCAGAAA GGGCGTCTTT	TCCTGCGGAA	S R R GTCTCGGAGA CAGAGCCTCT	F A Q TTCGCCCAGG AAGCGGGTCC
	A L P V CCCTGCCCGT GGGACGGGCA	TTGGGCGCGG	P D Y CCGGACTATA GGCCTGATAT	ACCCCCCCCT	V E T AGTGGAGACG TCACCTCTGC	W. K K TGGAAAAAGC ACCTTTTTCG	P D Y E CCGACTACGA GGCTGATGCT	P P V ACCACCTGTG TGGTGGACAC
	V H G GTCCATGGCT CAGGTACCGA	GCCCGCTTCC	P P K ACCTCCAAAG TGGAGGTTTC	S P P TCCCCTCCTG AGGGGAGGAC	TGCCTCCGCC	R K K TCGGAAGAAG AGCCTTCTTC	R T V V CGGACGGTGG GCCTGCCACC	TCCTCACTGA
5281 ———	ATCAACCCTA	TCTACTGCCT	L A E L TGGCCGAGCT ACCGGCTCGA	CGCCACCAGA	S F G S AGCTTTGGCA TCGAAACCGT	GCTCCTCAAC	S G I TTCCGGCATT AAGGCCGTAA	T G D ACGGGCGACA TGCCCGCTGT
	N T T T ATACGACAAC TATGCTGTTG	S S E ATCCTCTGAG TAGGAGACTC	P A P CCCGCCCTT	CTGGCTGCCC	P D S CCCCGACTCC GGGGCTGAGG	D A E GACGCTGAGT CTGCGACTCA	S Y S S CCTATTCCTC GGATAAGGAG	M P P CATGCCCCC GTACGGGGGG
+2	LEGI		P D L	S D G S	s w s T	v s s	EANA	E D V
5441							GAGGCCAACG CTCCGGTTGC	
5521 5521	CGTGTGCTGC	S M S) TCAATGTCTT AGTTACAGAA	ACTCTTGGAC	G A L AGGCGCACTC TCCGCGTGAG	GTCACCCCGT	GCGCCGCGA	E Q K AGAACAGAAA TCTTGTCTTT	L P I CTGCCCATCA GACGGGTAGT
				ACAATTTGGT	GTATTCCACC		S A C Q GTGCTTGCCA CACGAACGGT	
	K V T E AAAGTCACAT TTTCAGTGTA	TTGACAGACT	Q V L GCAAGTTCTG CGTTCAAGAC	D S H Y GACAGCCATT CTGTCGGTAA	ACCAGGACGT	ACTCAAGGAG	V K A A GTTAAAGCAG (CAATTTCGTC (A S K CGGCGTCAAA GCCGCAGTTT
5761	AGTGAAGGCT		CCGTAGAGGA			CACACTCAGC	K S K CAAATCCAAG GTTTAGGTTC	
				AGGCCGTAAC			K D L L AAGACCTTCT (TTCTGGAAGA (
		TAGACACTAC			TTTTCTGCGT		K G G R	GTAAGCCAGC
	CATIGIGGI	ATCTGTGATG	GTAGTACCGA	TICTIGCTCC	AAAAGACGCA	AGTCGGACTC	TTCCCCCCAG (ATTEGGTCG

pCMV-NS35

+2 6001	R L TCGTCT AGCAGA	~~~		F P TCCCC AGGGG	C 850	TCC	S V SCGT CGCA	GCGC	V GTGT CACA	GC -	GARA	K M AGATO	GG (CTTT	Y GTAC	CGA	CGTGC	V 1 STTAC	A i	K AAGC ITCG	TCC	CCT	_
+2	L A	V M	Ğ	s s	Y	G	F () Y	S	P	G	Q I	R		E (RI	. v	Q	A	W	ĸ	S	
6081	TGGCCG	TGAT ACTA	GGGA	AGCTC TCGAG	C TA	CGGA GCCT.	TTCC AAGG	AATA TTAT	CTCA GAGT	CC GG	AGGA TCCT	CAGC	GG (GTTG	AAT'	rcc	TCGT(GTT(SC (GTGG CACC	AAG TTC	TCC AGG	
+2 6161	K K AAGAAA TTCTTI	1000	C 3 3 7	G GGGGT CCCCA	T CT	S Y CGTA GCAT	TCAT	ACCC	R C GCTG GCGAC	CT	TTGA	S CTCC GAGG	AC .	AGTO	T ACT TGA	GAG	AGCG	O I ACATO IGTAO	CC	GTAC	GGA	E GGA CCT	
+2 6241	A I GGCAAT CCGTT	-	C2 24	C C GTTGT ACAACA	C 10	CTCC	D P ACCC TGGG	CCAA	A AGCCC	GC.	GTGG	A I CCAT GGTA	CA	AGTO	S L CCT GGA	CAC	CGAG.	R I AGGC TCCG	TT	Y TATG ATAC	TTG	GGG	_
+2 6321	G P GCCCTC CGGGAC			S F TTCAAC AAGTTC	~ ~~		n Aact TTGA	cccc	GATA	ירכ	CAGG	C TGCC ACGG	:GC	GCGZ	S AGCG TCGC	GCG	TACT	GACA	T AC TG	TAGO	C TGT	GGT	
+2 6401	N T AACACO TTGTG		OP.	C Y GCTACA CGATG1		K A AGGC	cccc	CCAC	A C GCCTC GGAC	TTC	GAGO	A CGCA	\GG	GCT	Q CCAG GGTC	GAC	TGCA	T M CCAT GGTA	GC	TCGT	GTO	G GTGG CACC	_
+2 6481	D CGACG	D L ACTTA IGAAT		V I GTTAT(CAATA(S A AGCGC	ccc	V GGTC(CCAG(~~~	GAGG	D P	CG	CGA	S I GCCT CGGA	GAG	AGCC	F TTCA AAGT	CG	GAG	A GCTA	ATGA	_
+2 6561	T R CCAGG GGTCC						P CCCCC GGGGG	CAC	Q P AACCI TTGG	202	ATAC	D CGAC1 CTG/	PTG	GAG	L CTC# GAG1	TAA	CATO	ATGC	S TC AG	CIC	N CAAC STTC	CGTG	
+2 6641	S V TCAGT AGTCA		-''	D G ACGGC TGCCG			K R AGAGO ICTCO		Y TACT ATGA	800	מחת -	r R CCCGT	4DT	CCC	T TACA ATG1	AACC	CCCC	L A	GA	GAG	CTG	A W CGTG GCAC	
+2 6721		T A CAGC! GTCG!		H T CACAC GTGTG	P TC C AG G		N S AATT(TTAA(L GCTA CGAT	ccc	מממ	I ATAA TATT.	TCA	TGT	F I	cccc	CAC	CTGT	GG	GCG	R AGG TCC	ATGA	
	I L TACTO			F TTTCT			L CCTTI GGAA'		A R CCAG GGTC	~~~	CCB	L GCTT CGAA	CAA	CAC	A GCCC CGG	CTCC	ATTO	C E CGAC CGCTC	GAT	CTA	CGG	A GGCC CCGG	· · · ·
6881	C Y TGCTA ACGAT			E P BAACCA CTTGGI			L P TACC ATGG		I ATCA TAGT	mmc		R L GACT CTGA	CCB	TCC	CCT GGA	CAGO	GCA	F :	CAC	TCC	ACA	S Y GTTA CAAT	
6961	-			I N NATCAA		V GGTG		a .mc					G GGG CCC	TAC	CGC	P 1 CCT'	r GCG.	A AGCT' TCGA	TGG	AGA	H CAC	CGGG	
7041	2 A R CCCGG GGGCG			R A GCGCTA CGCGA1			, A GGCC		G G GAGG CTC			A TGCC	מתאי	761	G TGGC ACCG	AAC	T ACC	L F TCTT AGAA	CAA	CTO	GGC	V AGTA TCAT	

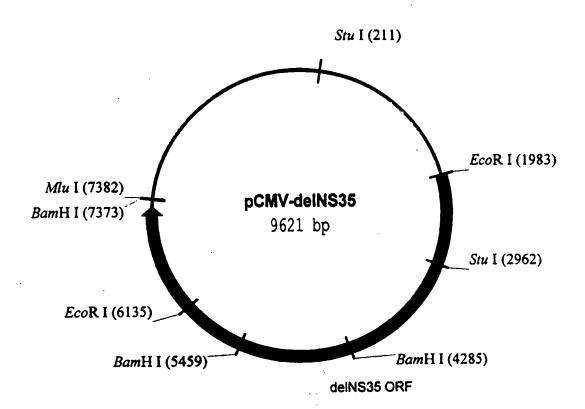
pCMV-NS35

+2 7121	R T K L AGAACAAAGC TCTTGTTTCG	TCABACTCAC	P I A TCCAATAGCG AGGTTATCGC	GCCGCTGGCC	AGCTGGACTT	S G W GTCCGGCTGG CAGGCCGACC	F T A G TTCACGGCTG AAGTGCCGAC	GCTACAGCGG
7201	CCCACACATT	Y H S \ TATCACAGCG ATAGTGTCGC	/ S H A TGTCTCATGC ACAGAGTACG	CCGGCCCCGC	TGGATCTGGT	TTTGCCTACT	L L A CCTGCTTGCT GGACGAACGA	A G V GCAGGGGTAG CGTCCCCATC
7281	G I Y L GCATCTACCT CGTAGATGGA	L P N CCTCCCCAAC GGAGGGGTTG	R CGATGAAGGT GCTACTTCCA	TGGGGTAAAC ACCCCATTTG	ACTCCGGCCT TGAGGCCGGA	AAAAAAAAA TTTTTTTTT	AAAAATCTAG TTTTTAGATC	AAAGGCGCGC TTTCCGCGCG
		BamHI	MluI					
7361	CAAGATATCA GTTCTATAGT	AGGATCCACT TCCTAGGTGA	ACGCGTTAGA TGCGCAATCT	GCTCGCTGAT CGAGCGACTA	CAGCCTCGAC GTCGGAGCTG	TGTGCCTTCT ACACGGAAGA	AGTTGCCAGC TCAACGGTCG	CATCTGTTGT GTAGACAACA
7441	TTGCCCCTCC AACGGGGAGG	CCCGTGCCTT GGGCACGGAA	CCTTGACCCT GGAACTGGGA	GGAAGGTGCC CCTTCCACGG	ACTCCCACTG TGAGGGTGAC	TCCTTTCCTA AGGAAAGGAT	ATAAAATGAG TATTTTACTC	GAAATTGCAT CTTTAACGTA
7521	CGCATTGTCT GCGTAACAGA	GAGTAGGTGT CTCATCCACA	CATTCTATTC GTAAGATAAG	TGGGGGGTGG ACCCCCCACC	GGTGGGGCAG CCACCCCGTC	GACAGCAAGG CTGTCGTTCC	GGGAGGATTG CCCTCCTAAC	GGAAGACAAT CCTTCTGTTA
7601	AGCAGGCATG TCGTCCGTAC	CTGGGGAGCT GACCCCTCGA	CTTCCGCTTC GAAGGCGAAG	CTCGCTCACT GAGCGAGTGA	GACTCGCTGC CTGAGCGACG	GCTCGGTCGT CGAGCCAGCA	TCGGCTGCGG AGCCGACGCC	CGAGCGGTAT GCTCGCCATA
7681	CAGCTCACTC GTCGAGTGAG	AAAGGCGGTA TTTCCGCCAT	ATACGGTTAT TATGCCAATA	CCACAGAATC GGTGTCTTAG	AGGGGATAAC TCCCCTATTG	GCAGGAAAGA CGTCCTTTCT	ACATGTGAGC TGTACACTCG	AAAAGGCCAG TTTTCCGGTC
7761	CAAAAGGCCA GTTTTCCGGT	GGAACCGTAA CCTTGGCATT	AAAGGCCGCG TTTCCGGCGC	TTGCTGGCGT AACGACCGCA	TTTTCCATAG AAAAGGTATC	GCTCCGCCCC	CCTGACGAGC GGACTGCTCG	ATCACAAAAA TAGTGTTTTT
7841	TCGACGCTCA AGCTGCGAGT	AGTCAGAGGT TCAGTCTCCA	GGCGAAACCC	GACAGGACTA CTGTCCTGAT	TAAAGATACC	AGGCGTTTCC TCCGCAAAGG	CCCTGGAAGC	TCCCTCGTGC AGGGAGCACG
7921	GCTCTCCTGT CGAGAGGACA	TCCGACCCTG AGGCTGGGAC	CCGCTTACCG GGCGAATGGC	GATACCTGTC CTATGGACAG	GCGGAAAGAG	CCTTCGGGAA GGAAGCCCTT	GCGTGGCGCT	TTCTCAATGC AAGAGTTACG
8001	TCACGCTGTA AGTGCGACAT	GGTATCTCAG CCATAGAGTC	TTCGGTGTAG AAGCCACATC	GTCGTTCGCT CAGCAAGCGA	CCAAGCTGGG GGTTCGACCC	CTGTGTGCAC GACACACGTG	GAACCCCCCG CTTGGGGGGC	TTCAGCCCGA AAGTCGGGCT
8081	CCGCTGCGCC	TTATCCGGTA AATAGGCCA1	A ACTATEGTET T TGATAGEAGA	TGAGTCCAAC ACTCAGGTTC	CCGGTAAGAC GGCCATTCTC	ACGACTTATO	GCCACTGGCA GCGGTGACCGT	GCAGCCACTG CGTCGGTGAC
8161	GTAACAGGAT CATTGTCCTA	TAGCAGAGCO	AGGTATGTAG TCCATACATC	GCGGTGCTAC	AGAGTTCTTC TCTCAAGAAC	AAGTGGTGGC	CTAACTACGG GATTGATGCC	CTACACTAGA GATGTGATCT
8241	AGGACAGTAT TCCTGTCATA	TTGGTATCT	GCGAGACGAC	AAGCCAGTTA TTCGGTCAAT	CCTTCGGAAA GGAAGCCTT	A AAGAGTTGGT	AGCTCTTGAT A TCGAGAACTA	CCGGCAAACA GGCCGTTTGT
8321	AACCACCGCT TTGGTGGCGA	GGTAGCGGTG CCATCGCCAG	GTTTTTTGT CAAAAAAACA	TTGCAAGCAC AACGTTCGTC	CAGATTACGO GTCTAATGCO	GCAGAAAAA GCGTCTTTTT	A AGGATCTCAA T TCCTAGAGTT	GAAGATCCTT CTTCTAGGAA
8401	TGATCTTTTC ACTAGAAAAC	TACGGGGTC ATGCCCCAG	I GACGCTCAGI A CTGCGAGTCA	GGAACGAAA CCTTGCTTT	A CTCACGTTA F GAGTGCAAT	A GGGATTTTGO I CCCTAAAACO	TCATGAGATT	ATCAAAAAGG TAGTTTTTCC

pCMV-NS35

8481	ATCTTCACCT	AGATCCTTTT	AAAATTAAA	TGAAGTTTTA	AATCAATCTA	AAGTATATAT	GAGTAAACTT	GGTCTGACAG
	, TAGAAGTGGA	TCTAGGAAAA	TTTTTAATTT	ACTTCAAAAT	TTAGTTAGAT	TTCATATATA	CTCATTTGAA	CCAGACTGTC
8561	TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGATC	TGTCTATTTC	GTTCATCCAT	AGTTGCCTGA	CTCCCGTCG
	AATGGTTACG	AATTAGTCAC	TCCGTGGATA	GAGTCGCTAG	ACAGATAAAG	CAAGTAGGTA	TCAACGGACT	GAGGGGCAGC
8641	TGTAGATAAC	TACGATACGG	GAGGGCTTAC	CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	GAGACCCACG	CTCACCGGCT
	ACATCTATTG	ATGCTATGCC	CTCCCGAATG	GTAGACCGGG	GTCACGACGT	TACTATGGCG	CTCTGGGTGC	GAGTGGCCGA
8721	CCAGATTTAT	CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG	AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	CCTCCATCCA
	GGTCTAAATA	GTCGTTATTT	GGTCGGTCGG	CCTTCCCGGC	TCGCGTCTTC	ACCAGGACGT	TGAAATAGGC	GGAGGTAGGT
8801	GTCTATTAAT	TGTTGCCGGG	AAGCTAGAGT	AAGTAGTTCG	CCAGTTAATA	GTTTGCGCAA	CGTTGTTGCC	ATTGCTACAG
	CAGATAATTA	ACAACGGCCC	TTCGATCTCA	TTCATCAAGC	GGTCAATTAT	CAAACGCGTT	GCAACAACGG	TAACGATGTC
8881	GCATCGTGGT	GTCACGCTCG	TCGTTTGGTA	TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT	CAAGGCGAGT	TACATGATCC
	CGTAGCACCA	CAGTGCGAGC	AGCAAACCAT	ACCGAAGTAA	GTCGAGGCCA	AGGGTTGCTA	GTTCCGCTCA	ATGTACTAGG
8961	CCCATGTTGT	GCAAAAAAGC	GGTTAGCTCC	TTCGGTCCTC	CGATCGTTGT	CAGAAGTAAG	TTGGCCGCAG	TGTTATCACT
	GGGTACAACA	CGTTTTTCG	CCAATCGAGG	AAGCCAGGAG	GCTAGCAACA	GTCTTCATTC	AACCGGCGTC	ACAATAGTGA
9041	CATGGTTATG	GCAGCACTGC	ATAATTCTCT	TACTGTCATG	CCATCCGTAA	GATGCTTTTC	TGTGACTGGT	GAGTACTCAA
	GTACCAATAC	CGTCGTGACG	TATTAAGAGA	ATGACAGTAC	GGTAGGCATT	CTACGAAAAG	ACACTGACCA	CTCATGAGTT
9121	CCAAGTCATT GGTTCAGTAA	CTGAGAATAG GACTCTTATC	TGTATGCGGC ACATACGCCG	GACCGAGTTG CTGGCTCAAC	CTCTTGCCCG	GCGTCAATAC CGCAGTTATG	GGGATAATAC CCCTATTATG	CGCGCCACAT GCGCGGTGTA
9201	AGCAGAACTT TCGTCTTGAA	TAAAAGTGCT ATTTTCACGA	CATCATTGGA GTAGTAACCT	AAACGTTCTT TTTGCAAGAA	CGGGGCGAAA	ACTCTCAAGG TGAGAGTTCC	ATCTTACCGC TAGAATGGCG	TGTTGAGATC ACAACTCTAG
9281	CAGTTCGATG GTCAAGCTAC	TAACCCACTC ATTGGGTGAG	GTGCACCCAA CACGTGGGTT	CTGATCTTCA GACTAGAAGT	GCATCTTTA CGTAGAAAAT	CTTTCACCAG GAAAGTGGTC	CGTTTCTGGG	TGAGCAAAA ACTCGTTTTT
9361	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	TAAGGGCGAC	ACGGAAATGT	TGAATACTCA	TACTCTTCCT	TTTTCAATAT
	GTCCTTCCGT	TTTACGGCGT	TTTTTCCCTT	ATTCCCGCTG	TGCCTTTACA	ACTTATGAGT	ATGAGAAGGA	AAAAGTTATA
9441	TATTGAAGCA ATAACTTCGT	TTTATCAGGG AAATAGTCCC	TTATTGTCTC	ATGAGCGGAT	ACATATTTGA TGTATAAACT	ATGTATTTAG TACATAAATC	AAAAATAAAC TTTTTATTTG	AAATAGGGGT TTTATCCCCA
9521	TCCGCGCACA	TTTCCCCGAA	AAGTGCCACC	TGACGTCTAP	GAAACCATTA	TTATCATGAC	ATTAACCTAT	AAAAATAGGC
	AGGCGCGTGT	AAAGGGGCTT	TTCACGGTGC	ACTGCAGATT	CTTTGGTAAT	AATAGTACTG	TAATTGGATA	TTTTTATCCG
9601	GTATCACGAG CATAGTGCTC	GCCCTTTCG1 CGGGAAAGCA						

FIGURE 4



pCMV-delNS35

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT
	AGCGCGCAAA	GCCACTACTG	CCACTTTTGG	AGACTGTGTA	CGTCGAGGGC	CTCTGCCAGT	GTCGAACAGA	CATTCGCCTA
81	GCCGGGAGCA	GACAAGCCCG	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA
	CGGCCCTCGT	CTGTTCGGGC	AGTCCCGCGC	AGTCGCCCAC	AACCGCCCAC	AGCCCCGACC	GAATTGATAC	GCCGTAGTCT
					Sti	ıI		
161	GCAGATTGTA	CTGAGAGTGC	ACCATATGAA	GCTTTTTGCA	AAAGCCTAGG	CCTCCAAAAA	AGCCTCCTCA	CTACTTCTGG
	CGTCTAACAT	GACTCTCACG	TGGTATACTT	CGAAAAACGT	TTTCGGATCC	GGAGGTTTTT	TCGGAGGAGT	GATGAAGACC
241	AATAGCTCAG TTATCGAGTC		GGCCTCGGCC CCGGAGCCGG					
321	ACTGGGCGGG TGACCCGCCC		TIGGCTATTG AACCGATAAC					
401	CATGTCCAAT GTACAGGTTA		TGTTGACATT ACAACTGTAA					
481	AGCCCATATA TCGGGTATAT		CGTTACATAA GCAATGTATT					
561	GACGTCAATA CTGCAGTTAT		TTCCCATAGT AAGGGTATCA					
641	AAACTGCCCA TTTGACGGGT		CATCAAGTGT GTAGTTCACA					
721	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	CGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC
	CGGACCGTAA	TACGGGTCAT	GTACTGGAAT	GCCCTGAAAG	GATGAACCGT	CATGTAGATG	CATAATCAGT	AGCGATAATG
801	CATGGTGATG GTACCACTAC		AGTACACCAA TCATGTGGTT					
881	TTGACGTCAA AACTGCAGTT		TTTTGGCACC AAAACCGTGG					
961	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG
	GTTTACCCGC	CATCCGCACA	TGCCACCCTC	CAGATATATT	CGTCTCGAGC	AAATCACTTG	GCAGTCTAGC	GGACCTCTGC
.041	CCATCCACGC	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	GGAACGGTGC	ATTGGAACGC
	GGTAGGTGCG	ACAAAACTGG	AGGTATCTTC	TGTGGCCCTG	GCTAGGTCGG	AGGCGCCGGC	CCTTGCCACG	TAACCTTGCG
121	GGATTCCCCG	TGCCAAGAGT	GACGTAAGTA	CCGCCTATAG	ACTCTATAGG	CACACCCCTT	TGGCTCTTAT	GCATGCTATA
	CCTAAGGGGC	ACGGTTCTCA	CTGCATTCAT	GGCGGATATC	TGAGATATCC	GTGTGGGGAA	ACCGAGAATA	CGTACGATAT
201	CTGTTTTTGG	CTTGGGGCCT	ATACACCCCC	GCTCCTTATG	CTATAGGTGA	TGGTATAGCT	TAGCCTATAG	GTGTGGGTTA
	GACAAAAACC	GAACCCCGGA	TATGTGGGGG	CGAGGAATAC	GATATCCACT	ACCATATCGA	ATCGGATATC	CACACCCAAT
1281	TTGACCATTA	TTGACCACTC	CCCTATTGGT	GACGATACTT	TCCATTACTA	ATCCATAACA	TGGCTCTTTG	CCACAACTAT
	AACTGGTAAT	AACTGGTGAG	GGGATAACCA	CTGCTATGAA	AGGTAATGAT	TAGGTATTGT	ACCGAGAAAC	GGTGTTGATA
1361	CTCTATTGGC	TATATGCCAA	TACTCTGTCC	TTCAGAGACT	GACACGGACT	CTGTATTTT	ACAGGATGGG	GTCCATTTAT
	GAGATAACCG	ATATACGGTT	ATGAGACAGG	AAGTCTCTGA	CTGTGCCTGA	GACATAAAA	TGTCCTACCC	CAGGTAAATA

1441							TAGCGTGGGA ATCGCACCCT	
1521	CTCGGGTACG GAGCCCATGC	TGTTCCGGAC ACAAGGCCTG	ATGGGCTCTT TACCCGAGAA	CTCCGGTAGC GAGGCCATCG	GGCGGAGCTT CCGCCTCGAA	CCACATCCGA GGTGTAGGCT	GCCCTGGTCC CGGGACCAGG	CATCCGTCCA GTAGGCAGGT
1601							GCACAATGCC CGTGTTACGG	
1681							GCTCGCACCT CGAGCGTGGA	
1761	GGAAGACTTA CCTTCTGAAT	AGGCAGCGGC TCCGTCGCCG	AGAAGAAGAT TCTTCTTCTA	GCAGGCAGCT CGTCCGTCGA	GAGTTGTTGT CTCAACAACA	ATTCTGATAA TAAGACTATT	GAGTCAGAGG CTCAGTCTCC	TAACTCCCGT ATTGAGGGCA
1841	TGCGGTGCTG ACGCCACGAC	TTAACGGTGG AATTGCCACC	AGGGCAGTGT TCCCGTCACA	AGTCTGAGCA TCAGACTCGT	GTACTCGTTG CATGAGCAAC	CTGCCGCGCG GACGGCGCGC	CGCCACCAGA GCGGTGGTCT	CATAATAGCT GTATTATCGA
+2							EcoRI	M A A
1921	GACAGACTAA CTGTCTGATT	CAGACTGTTC GTCTGACAAG	CTTTCCATGG GAAAGGTACC	GTCTTTTCTG CAGAAAAGAC	CAGTCACCGT GTCAGTGGCA	CGTCGACCTA GCAGCTGGAT	AGAATTCACC TCTTAAGTGG	ATGGCTGCAT TACCGACGTA
2001	Y A A Q ATGCAGCTCA TACGTCGAGT	GGGCTATAAG	V L V GTGCTAGTAC CACGATCATG	TCAACCCCTC	TGTTGCTGCA	T L G ACACTGGGCT TGTGACCCGA	F G A Y TTGGTGCTTA AACCACGAAT	CATGTCCAAG
+2 2081	A H G GCTCATGGGA CGAGTACCCT	TCGATCCTAA	CATCAGGACC	G V R GGGGTGAGAA CCCCACTCTT	CAATTACCAC	TGGCAGCCCC	I T Y ATCACGTACT TAGTGCATGA	S T Y G CCACCTACGG GGTGGATGCC
+2	CAAGTTCCTT	A D G GCCGACGGCG CGGCTGCCGC	GGTGCTCGGG	GGGCGCTTAT	GACATAATAA	TTTGTGACGA	C H S GTGCCACTCC CACGGTGAGG	T D A ACGGATGCCA TGCCTACGGT
+2	T S I L CATCCATCTT GTAGGTAGAA	GGGCATTGGC	T V L ACTGTCCTTG TGACAGGAAC	ACCAAGCAGA	GACTGCGGGG	GCGAGACTGG	V V L A TTGTGCTCGC AACACGAGCG	CACCGCCACC
+2 2321	P P G CCTCCGGGCT GGAGGCCCGA	CCGTCACTGT	GCCCCATCCC	N I E AACATCGAGG TTGTAGCTCC	AGGTTGCTCT	GTCCACCACC	G E I GGAGAGATCC CCTCTCTAGG	CTTTTTACGG
+2 2401	CAACGCTATC	P L E CCCCTCGAAG GGGGAGCTTC	TAATCAAGGG	GGGGAGACAT	CTCATCTTCT	GTCATTCAAA	K K C GAAGAAGTGC CTTCTTCACG	D E L GACGAACTCG CTGCTTGAGC
+2	A A K L CCGCAAAGCT GGCGTTTCGA	CCTCCCATTC	CCCATCAATG	A V A Y CCGTGGCCTA GGCACCGGAT	CTACCGCGGT	L D V CTTGACGTGT GAACTGCACA	S V I P CCGTCATCCC GGCAGTAGGG	GACCAGCGGC
+2 2561	D V V GATGTTGTCG CTACAACAGO	TOGTGGCAAC	CGATGCCCTC	M T G ATGACCGGCT TACTGGCCGA	ATACCGGCGA	CTTCGACTCG	V I D GTGATAGACT CACTATCTGA	GCAATACGTG

2641	V T Q TGTCACCCAG ACAGTGGGTC	ACAGTCGATT	S L D TCAGCCTTGA AGTCGGAACT	CCCTACCTTC	ACCATTGAGA	I T L CAATCACGCT GTTAGTGCGA	CCCCCAAGAT	A V S GCTGTCTCCC CGACAGAGGG
2721	R T Q R GCACTCAACG CGTGAGTTGC	TOGGGGCAGG	T G R (ACTGGCAGGG TGACCGTCCC	GGAAGCCAGG	CATCTACAGA	F V A P TTTGTGGCAC AAACACCGTG	G E R CGGGGGAGCG GCCCCCTCGC	P S G CCCCTCCGGC GGGGAGGCCG
+2 2801	M F D S ATGTTCGACT TACAAGCTGA	CGTCCGTCCT	CTGTGAGTGC	Y D A C TATGACGCAG ATACTGCGTC	GCTGTGCTTG	Y E L GTATGAGCTC CATACTCGAG	T P A E ACGCCCGCCG TGCGGGCGGC	AGACTACAGT
+2	RLR	A Y M	N T P G	L P V	CQDI	t "L E F	W E G	V F T StuI
2881	TAGGCTACGA ATCCGATGCT	GCGTACATGA CGCATGTACT	ACACCCCGGG TGTGGGGCCC	GCTTCCCGTG CGAAGGGCAC	TGCCAGGACC ACGGTCCTGG	ATCTTGAATT TAGAACTTAA	TTGGGAGGGC AACCCTCCCG	GTCTTTACAG CAGAAATGTC
+2	G L T H StuI	I D A	H F L	S Q T K	Q S G	ENLE	YLV	A Y Q
2961	CCCTCACTCA	TATAGATGCC ATATCTACGG	CACTTTCTAT GTGAAAGATA	CCCAGACAAA GGGTCTGTTT	GCAGAGTGGG CGTCTCACCC	GAGAACCTTC CTCTTGGAAG	CTTACCTGGT GAATGGACCA	AGCGTACCAA TCGCATGGTT
+2 3041	A T V GCCACCGTGT CGGTGGCACA	GCGCTAGGGC	TCAACCCCCT	P P S CCCCCATCGT GGGGGTAGCA	GGGACCAGAT	W K C GTGGAAGTGT CACCTTCACA	L I R I TTGATTCGCC AACTAAGCGG	TCAAGCCCAC
+2	CCTCCATGGG	P T P CCAACACCCC GGTTGTGGGG	TGCTATACAG	ACTGGGCGCT	V Q N GTTCAGAATG CAAGTCTTAC	E I T L AAATCACCCT TTTAGTGGGA	T H P GACGCACCCA CTGCGTGGGT	V T K GTCACCAAAT CAGTGGTTTA
+2 3201	Y I M T ACATCATGAC TGTAGTACTG	ATCCATCTCC	A D L GCCGACCTGG CGGCTGGACC	AGGTCGTCAC	GAGCACCTGG	V L V C GTGCTCGTTG CACGAGCAAC	GCGGCGTCCT	A A L GGCTGCTTTG CCGACGAAAC
+2 3281	A A Y GCCGCGTATT CGGCGCATAA	CCCTCTCAAC	AGGCTGCGTG	GTCATAGTGG	GCAGGGTCGT	L S G CTTGTCCGGG GAACAGGCCC	K P A AAGCCGGCAA	TCATACCTGA
+2	CACCCAACTC	L Y R CTCTACCGAG GAGATGGCTC	ACTTCCATCA	GATGGAAGAG	C S Q TGCTCTCAGO ACGAGAGTCG	H L P Y ACTTACCGTA TGAATGGCAT	CATCGAGCAA	G M M GGGATGATGC CCCTACTACG
+2	L A E C TCGCCGAGCA AGCGGCTCGT	CTTCNACCAC	K A L AAGGCCCTCC TTCCGGGAGC	CCCTCCTGC	GACCGCGTCC	R Q A COTCAGGCAG GCAGTCCGTC	AGGTTATCGC	P A V CCCTGCTGTC GGGACGACAG
+2 3521	Q T N CAGACCAACT GTCTGGTTGA		CC3C3CCCTT(TOCCCGAAGO	ATATGTGGAA	F I S CTTCATCAGT GAAGTAGTCA	G I Q GGGATACAAT CCCTATGTTA	ACTTGGCGGG
+2 3601	CHACACA ACC		ACCCCCCCA	TOTTOTTO	M A F ATGGCTTTTA TACCGAAAAT	T A A V A CAGCTGCTGT C GTCGACGACA	CACCAGCCCA	L T T CTAACCACTA .GATTGGTGAT
3681	S Q T I GCCAAACCCT CGGTTTGGG		3 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	COMPORTEC	TOCCCACCTO	A A P GCCGCCCCG GCGGGGGGG	GTGCCGCTAC	TGCCTTTGTG

3761	G A G GGCGCTGG CCGCGACC														
3841	GGGCGTGGG CCCGCACCG	G GGAC	CTCTTC	V A TGGCAT ACCGTA	TCAA	GATCA	TCACC	CCTCA	CCTCC	P S CCTCC GGAGG	B CCCB	CC3.CC	L V TGGTC ACCAG		
	P A I CCGCCATCO GGCGGTAGO	T CTCC	P G GCCCGGA GGGCCT	GCCCTC	STAG	TCGGC	CTCCT	C CTGTG GACAC	CACCA	3 T 3 C T	R GCGCC CGCGG		V G GTTGG CAACC		G E GCGAG CGCTC
4001	G A V GGGGCAGTC CCCCGTCAC	CAGTO	M N GATGAA CTACTT	CCGGCT	I GATA CTAT	GCCTT	A CGCCT GCGGA	CCCGG	G N GGGAA CCCTT	CCATG	V S TTTCC AAAGG	CCCAC	CCRCT	Y ACGTG TGCAC	~~~~
4081	S D A GAGCGATGO CTCGCTACO	A GCTG	CCCGCG	V T TCACTGO AGTGACO	CCAT	ACTCA	GCAGC	CTCAC	TCTAA	CCCAG	CTCCT	CACCC	R L GACTG CTGAC	H Q CACCA GTGGT	~~~~
	I S S TAAGCTCGC ATTCGAGCC	A GTGT	T T ACCACT TGGTGA	CCATGC	CCG	GTTCC	TGGCT	AAGGG	D I ACATC IGTAG	TOGGA	CTCCA	I C TATGC: ATACG:	TACCT	CTTC	S D GCGAC CGCTG
+2	F K T	W L	K A	K L	M	P Q	L i		I P	F	v s	C Q	R C	; Y	K G
4241	TTTAAGACC	T GGCT	AAAAGC TTTTCG	TAAGCTO	CATG	CCACA GGTGT	GCTGC CGACG	CTGGG	ATCCC PAGGG	CTTTG GAAAC	IGTCC ACAGG	TGCCAC	GCGC GCGC	GGTAT.	AAGGG TTCCC
4321	V W R GGTCTGGCG CGAGACCGC	A GGGG	D G ACGGCA TGCCGT	TCATGC	T ACAC TGTG	TCGCT	C H GCCAC CGGTG	TGTGG	A E AGCTG CCGAC	AGATO	T G ACTGG IGACC	ACATO	/ K CCAAA AGTTT	N G AACGGG TTGCC	~3.003
	M R I TGAGGATCG ACTCCTAGC	T CGGT	P R CCTAGG GGATCC	T C ACCTGCA TGGACG1	GGA .	ACATG'	TGGAG	G 1 TGGGAC ACCCTC	CTTC	CCCATT	N A TAATG ATTAC	CCTACE	T T ACCAC TGGTG	ccccc	P C CCTGT GGACA
	T P L ACCCCCTT TGGGGGGAA	C CTGC	P N GCCGAA CGGCTT	CTACACO	TTC	GCGCT	ATGGA	V GGGTG1 CCCAC	CTGC	AGAGG	E Y NATAC ITATG	GTGGAC	I R SATAA STATT	GGCAGG	V G GTGGG CACCC
+2 4561	D F H GGACTTCCA CCTGAAGGT	C TACG	V T (TGACGG ACTGCC	GTATGAC	TAC	TGACA	n L ATCTT FAGAA	AAATGO	P C CCGT GGCA	GCCAGO	V P STCCC CAGGG	S E ATCGCC TAGCGC	CGAA	F F TTTTTC AAAAA	JACAC .
	E L D AATTGGACG TTAACCTGC	G GGTG	R L CGCCTA GCGGAT	H R CATAGGI GTATCCA	TTG (CGCCCC	P C CCCTG GGGAC	CAAGCO	L CTTG GAAC	CTGCGC	E E GAGG CTCC	AGGTAT	S F CATT GTAA	CAGAGT	/ G PAGGA NTCCT
4721	L H E CTCCACGAA GAGGTGCTT	r ACCC	JURITU	GTCGCAA	TTA (CCTTG	CGAGC	CCGAAC	CGGA	CGTGGC	CGTG	TTGACG	TCCA	TGCTC	CTGA
4801	P S H TCCCTCCCA AGGGAGGGT	r ataa	CAGCAG	AGGCGGC	CGG (GCGAAC	GTTG	GCGAGG	GGAT	CACCCC	CCTC	TGTGGC	CAGC	TCCTCC	GCTA

+2 S Q L S A P S L K A T C T A N H D S P D A E L I E A N 4881 GCCAGCTATC CGCTCCATCT CTCAAGGCAA CTTGCACCGC TAACCATGAC TCCCCTGATG CTGAGCTCAT AGAGGCCAAC CGGTCGATAG GCGAGGTAGA GAGTTCCGTT GAACGTGGCG ATTGGTACTG AGGGGACTAC GACTCGAGTA TCTCCGGGTTC	; ;
+2 L L W R · Q E M G G N I T R V E S E N K V V I L D S F E A M C CTCCTATGGA GGCAGGAGAT GGGCGGCAAC ATCACCAGGG TTGAGTCAGA AAACAAAGTG GTGATTCTGG ACTCCTTCGAGAGGAAGCT GAGGATACCT CCGTCCTCTA CCCGCCGTTG TAGTGGTCCC AACTCAGTCT TTTGTTTCAC CACTAAGACC TGAGGAAGCT	
+2 P L V A E E D E R E I S V P A E I L R K S R R F A Q 5041 TCCGCTTGTG GCGGAGGGG ACGAGGGGG GATCTCCGTA CCCGCAGAAA TCCTGCGGAA GTCTCGGAGA TTCGCCCAGGAGCGCGAACAC CGCCTCCTCC TGCTCGCCCT CTAGAGGCAT GGGCGTCTTT AGGACGCCTT CAGAGCCTCT AAGCGGGTCC	
+2 A L P V W A R P D Y N P P L V E T W K K P D Y E P P V 5121 CCCTGCCGT TTGGGCGCG CCGGACTATA ACCCCCCGCT AGTGGAGACG TGGAAAAAGC CCGACTACGA ACCACCTGTG GGGACGGCA AACCCGCGCC GGCCTGATAT TGGGGGGCGA TCACCTCTGC ACCTTTTCG GGCTGATGCT TGGTGGACAC	 :
+2 V H G C P L P P P K S P P V P P P R K K R T V V L T E 5201 GTCCATGGCT GCCCGCTTCC ACCTCCAAAG TCCCCTCCTG TGCCTCCGCC TCGGAAGAAG CGGACGGTGG TCCTCACTGA CAGGTACCGA CGGGCGAAGG TGGAGGTTC AGGGGAGGAC ACGGAGGCGG AGCCTTCTTC GCCTGCCACC AGGAGTGACT	
+2 S T L S T A L A E L A T R S F G S S S T S G I T G D 5281 ATCAACCCTA TCTACTGCCT TGGCCGAGCT CGCCACCAGA AGCTTTGGCA GCTCCTCAAC TTCCGGCATT ACGGGCGACA TAGTTGGGAT AGATGACGGA ACCGGCTCGA GCGGTGGTCT TCGAAACCGT CGAGGAGTTG AAGGCCGTAA TGCCCGCTGT	
+2 N T T T S S E P A P S G C P P D S D A E S Y S S M P P 5361 ATACGACAAC ATCCTCTGAG CCCGCCCTT CTGGCTGCCC CCCGACTCC GACGCTGAGT CCTATTCCTC CATGCCCCCC TATGCTGTTG TAGGAGACTC GGGCGGGGAA GACCGACGGG GGGCTGAGG CTGCGACTCA GGATAAGGAG GTACGGGGGG	
+2 LEGE PGD POL SDGS WST VSS EANA EDV BamHI	
5441 CTGGAGGGGG AGCCTGGGGA TCCGGATCTT AGCGACGGGT CATGGTCAAC GGTCAGTAGT GAGGCCAACG CGGAGGATGT GACCTCCCCC TCGGACCCCT AGGCCTAGAA TCGCTGCCCA GTACCAGTTG CCAGTCATCA CTCCGGTTGC GCCTCCTACA	
+2 V C C S M S Y S W T G A L V T P C A A E E Q K L P I 5521 CGTGTGCTGC TCAATGTCTT ACTCTTGGAC AGGCGCACTC GTCACCCCGT GCGCCGCGGA AGAACAGAAA CTGCCCATCA GCACACGACG AGTTACAGAA TGAGAACCTG TCCGCGTGAG CAGTGGGGCA CGCGGCGCCT TCTTGTCTTT GACGGGTAGT	
+2 N A L S N S L L R H H N L V Y S T T S R S A C Q R Q K 5601 ATGCACTAAG CAACTCGTTG CTACGTCACC ACAATTTGGT GTATTCCACC ACCTCACGCA GTGCTTGCCA AAGGCAGAAG TACGTGATTC GTTGAGCAAC GATGCAGTGG TGTTAAACCA CATAAGGTGG TGGAGTGCGT CACGAACGGT TTCCGTCTTC	
+2 K V T F D R L Q V L D S H Y Q D V L K E V K A A A S K 5681 AAAGTCACAT TTGACAGACT GCAAGTTCTG GACAGCCATT ACCAGGACGT ACTCAAGGAG GTTAAAGCAG CGGCGTCAAA TTTCAGTGTA AACTGTCTGA CGTTCAAGAC CTGTCGGTAA TGGTCCTGCA TGAGTTCCTC CAATTTCGTC GCCGCAGTTT	
+2 V K A N L L S V E E A C S L T P P H S A K S K F G Y 5761 AGTGAAGGCT AACTTGCTAT CCGTAGAGGA AGCTTGCAGC CTGACGCCCC CACACTCAGC CAAATCCAAG TTTGGTTATG TCACTTCCGA TTGAACGATA GGCATCTCCT TCGAACGTCG GACTGCGGGG GTGTGAGTCG GTTTAGGTTC AAACCAATAC	
+2 G A K D V R C H A R K A V T H I N S V W K D L L E D N 5841 GGGCAAAAGA CGTCCGTTGC CATGCCAGAA AGGCCGTAAC CCACATCAAC TCCGTGTGGA AAGACCTTCT GGAAGACAAT CCCGTTTTCT GCAGGCAACG GTACGGTCTT TCCGGCATTG GGTGTAGTTG AGGCACACCT TTCTGGAAGA CCTTCTGTTA	_
+2 V T P I D T T I M A K N E V F C V Q P E K G G R K P A 5921 GTAACACCAA TAGACATAC CATCATGGCT AAGAACGAGG TTTTCTGCGT TCAGCCTGAG AAGGGGGGTC GTAAGCCAGC CATTGTGGTT ATCTGTGATG GTAGTACCGA TTCTTGCTCC AAAAGACGCA AGTCGGACTC TTCCCCCCAG CATTCGGTCG	_

pCMV-delNS35

6001	TCGTCTCATC	V F P GTGTTCCCCG CACAAGGGGC	ATCTGGGCGT	GCGCGTGTGC	E K M . GAAAAGATGG CTTTTCTACC	CTTTGTACGA	V V T CGTGGTTACA GCACCAATGT	K L P AAGCTCCCCT TTCGAGGGGA
+2	LAVM	. G S S	Y G F	Q Y S P	G Q R	V E F I	A D V	w K S
6081	TGGCCGTGAT ACCGGCACTA	GGGAAGCTCC CCCTTCGAGG	TACGGATTCC ATGCCTAAGG	AATACTCACC TTATGAGTGG	AGGACAGCGG TCCTGTCGCC	GTTGAATTCC	TCGTGCAAGC AGCACGTTCG	GTGGAAGTCC CACCTTCAGG
	K K T AAGAAAACCC TTCTTTTGGG	P M G F CAATGGGGTT GTTACCCCAA	S Y D CTCGTATGAT GAGCATACTA	T R C ACCCGCTGCT TGGGCGACGA	TTGACTCCAC	V T E AGTCACTGAG TCAGTGACTC	S D I F AGCGACATCC TCGCTGTAGG	GTACGGAGGA
6241	GGCAATCTAC	Q C C CAATGTTGTG GTTACAACAC	ACCTCGACCC	Q A R CCAAGCCCGC GGTTCGGGCG	V A I GTGGCCATCA CACCGGTAGT	AGTCCCTCAC	E R L CGAGAGGCTT GCTCTCCGAA	Y V G TATGTTGGGG ATACAACCCC
	G P L T GCCCTCTTAC CGGGAGAATG	N S R CAATTCAAGG GTTAAGTTCC	G E N G GGGGAGAACT CCCCTCTTGA	GCGGCTATCG	R C R CAGGTGCCGC GTCCACGGCG	A S G V GCGAGCGGCG CGCTCGCCGC	TACTGACAAC	S C G TAGCTGTGGT ATCGACACCA
-		T C Y I CTTGCTACAT GAACGATGTA	CAAGGCCCGG		GAGCCGCAGG	GCTCCAGGAC		TCGTGTGTGG
6481	CGACGACTTA	V V I GTCGTTATCT CAGCAATAGA	GTGAAAGCGC	GGGGGTCCAG		CGAGCCTGAG	AGCCTTCACG	
		A P P CGCCCCCCT GCGGGGGGA		CACAACCAGA	ATACGACTTG		CATCATGCTC	
		H D G A ACGACGCGC TGCTGCCGCG	TGGAAAGAGG		TCACCCGTGA	CCCTACAACC		GAGCTGCGTG
+2 6721	GGAGACAGCA	R H T AGACACACTC TCTGTGTGAG		CTGGCTAGGC		TGTTTGCCCC	CACACTGTGG	
-		H F F CCATTTCTTT GGTAAAGAAA	AGCGTCCTTA		CCAGCTTGAA	CAGGCCCTCG		CTACGGGGCC
		I E P L TAGAACCACT ATCTTGGTGA		CCAATCATTC	AAAGACTCCA	TGGCCTCAGC		TCCACAGTTA
+2 6961		E I N I GAAATCAATA CTTTAGTTAT	GGGTGGCCGC	ATGCCTCAGA		TACCGCCCTT		
		R A R CCGCGCTAGG GGCGCGATCC		GAGGAGGCAG	GGCTGCCATA		ACCTCTTCAA	
								

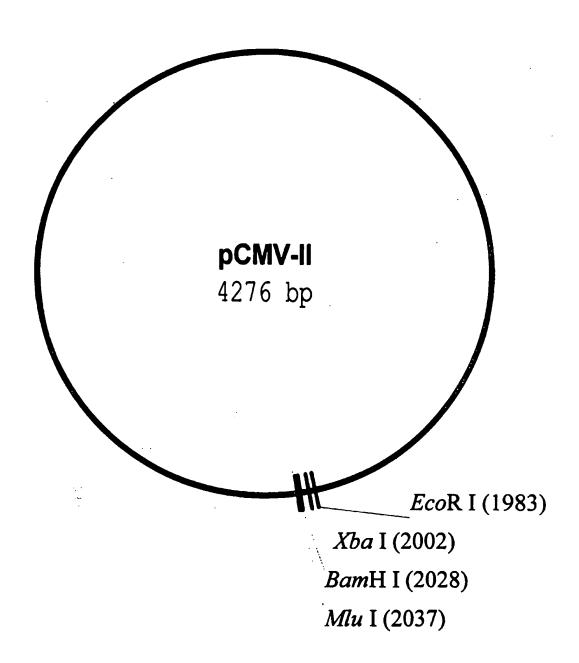
pCMV-delNS35

	R T K I AGAACAAAGC TCTTGTTTCG	TCAAACTCAC			AGCTGGACTT			GCTACAGCGG
+2 7201	GGGAGACATT			CCGGCCCGC		TTTGCCTACT		
	G I Y L GCATCTACCT CGTAGATGGA		R CGATGAAGGT GCTACTTCCA					
		BamHI	MluI					
7361	CAAGATATCA GTTCTATAGT		ACGCGTTAGA TGCGCAATCT					
7441	TTGCCCCTCC AACGGGGAGG		CCTTGACCCT GGAACTGGGA					
7521	CGCATTGTCT	GAGTAGGTGT	CATTĈTATTC	TGGGGGGTGG	GGTGGGGCAG	GACAGCAAGG	GGGAGGATTG	GGAAGACAAT
	GCGTAACAGA	CTCATCCACA	GTAAGATAAG	ACCCCCCACC	CCACCCCGTC	CTGTCGTTCC	CCCTCCTAAC	CCTTCTGTTA
7601	AGCAGGCATG	CTGGGGAGCT	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	GCTCGGTCGT	TCGGCTGCGG	CGAGCGGTAT
	TCGTCCGTAC	GACCCCTCGA	GAAGGCGAAG	GAGCGAGTGA	CTGAGCGACG	CGAGCCAGCA	AGCCGACGCC	GCTCGCCATA
7681	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG
	GTCGAGTGAG	TTTCCGCCAT	TATGCCAATA	GGTGTCTTAG	TCCCCTATTG	CGTCCTTTCT	TGTACACTCG	TTTTCCGGTC
7761			AAAGGCCGCG TTTCCGGCGC					
7841	TCGACGCTCA	AGTCAGAGGT	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC
	AGCTGCGAGT	TCAGTCTCCA	CCGCTTTGGG	CTGTCCTGAT	ATTTCTATGG	TCCGCAAAGG	GGGACCTTCG	AGGGAGCACG
7921	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	GATACCTGTC	CGCCTTTCTC	CCTTCGGGAA	GCGTGGCGCT	TTCTCAATGC
	CGAGAGGACA	AGGCTGGGAC	GGCGAATGGC	CTATGGACAG	GCGGAAAGAG	GGAAGCCCTT	CGCACCGCGA	AAGAGTTACG
8001	TCACGCTGTA	GGTATCTCAG	TTCGGTGTAG	GTCGTTCGCT	CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	TTCAGCCCGA
	AGTGCGACAT	CCATAGAGTC	AAGCCACATC	CAGCAAGCGA	GGTTCGACCC	GACACACGTG	CTTGGGGGGC	AAGTCGGGCT
8081	CCGCTGCGCC	TTATCCGGTA	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG
	GGCGACGCGG	AATAGGCCAT	TGATAGCAGA	ACTCAGGTTG	GGCCATTCTG	TGCTGAATAG	CGGTGACCGT	CGTCGGTGAC
8161	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA
	CATTGTCCTA	ATCGTCTCGC	TCCATACATC	CGCCACGATG	TCTCAAGAAC	TTCACCACCG	GATTGATGCC	GATGTGATCT
8241	AGGACAGTAT	TTGGTATCTG	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	CCGGCAAACA
	TCCTGTCATA	AACCATAGAC	GCGAGACGAC	TTCGGTCAAT	GGAAGCCTTT	TTCTCAACCA	TCGAGAACTA	GGCCGTTTGT
8321	AACCACCGCT	GGTAGCGGTG	GTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	GCAGAAAAA	AGGATCTCAA	GAAGATCCTT
	TTGGTGGCGA	CCATCGCCAC	CAAAAAAACA	AACGTTCGTC	GTCTAATGCG	CGTCTTTTT	TCCTAGAGTT	CTTCTAGGAA
8401	TGATCTTTTC	TACGGGGTCT	GACGCTCAGT	GGAACGAAAA	CTCACGTTAA	GGGATTTTGG	TCATGAGATT	ATCAAAAAGG
	ACTAGAAAAG	ATGCCCCAGA	CTGCGAGTCA	CCTTGCTTTT	GAGTGCAATT	CCCTAAAACC	AGTACTCTAA	TAGTTTTTCC

pCMV-delNS35

ATCTTCACCT	AGATCCTTTT	AAATTAAAA	TGAAGTTTTA	AATCAATCTA	AAGTATATAT	GAGTAAACTT	GGTCTGACAG
TAGAAGTGGA	TCTAGGAAAA	TTTTAATTT	ACTTCAAAAT	TTAGTTAGAT	TTCATATATA	CTCATTTGAA	CCAGACTGTC
TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGATC	TGTCTATTTC	GTTCATCCAT	AGTTGCCTGA	CTCCCCGTCG
AATGGTTACG	AATTAGTCAC	TCCGTGGATA	GAGTCGCTAG	ACAGATAAAG	CAAGTAGGTA	TCAACGGACT	GAGGGGCAGC
TGTAGATAAC	TACGATACGG	GAGGGCTTAC	CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	GAGACCCACG	CTCACCGGCT
ACATCTATTG	ATGCTATGCC	CTCCCGAATG	GTAGACCGGG	GTCACGACGT	TACTATGGCG	CTCTGGGTGC	GAGTGGCCGA
CCAGATTTAT	CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG	AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	CCTCCATCCA
GGTCTAAATA	GTCGTTATTT	GGTCGGTCGG	CCTTCCCGGC	TCGCGTCTTC	ACCAGGACGT	TGAAATAGGC	GGAGGTAGGT
GTCTATTAAT	TGTTGCCGGG	AAGCTAGAGT	AAGTAGTTCG	CCAGTTAATA	GTTTGCGCAA	CGTTGTTGCC	ATTGCTACAG
CAGATAATTA	ACAACGGCCC	TTCGATCTCA	TTCATCAAGC	GGTCAATTAT	CAAACGCGTT	GCAACAACGG	TAACGATGTC
GCATCGTGGT	GTCACGCTCG	TCGTTTGGTA	TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT	CAAGGCGAGT	TACATGATCC
CGTAGCACCA	CAGTGCGAGC	AGCAAACCAT	ACCGAAGTAA	GTCGAGGCCA	AGGGTTGCTA	GTTCCGCTCA	ATGTACTAGG
CCCATGTTGT	GCAAAAAAGC	GGTTAGCTCC	TTCGGTCCTC	CGATCGTTGT	CAGAAGTAAG	TTGGCCGCAG	TGTTATCACT
GGGTACAACA	CGTTTTTTCG	CCAATCGAGG	AAGCCAGGAG	GCTAGCAACA	GTCTTCATTC	AACCGGCGTC	ACAATAGTGA
CATGGTTATG	GCAGCACTGC	ATAATTCTCT	TACTGTCATG	CCATCCGTAA	GATGCTTTTC	TGTGACTGGT	GAGTACTCAA
GTACCAATAC	CGTCGTGACG	TATTAAGAGA	ATGACAGTAC	GGTAGGCATT	CTACGAAAAG	ACACTGACCA	CTCATGAGTT
CCAAGTCATT	CTGAGAATAG	TGTATGCGGC	GACCGAGTTG	CTCTTGCCCG	GCGTCAATAC	GGGATAATAC	CGCGCCACAT
GGTTCAGTAA	GACTCTTATC	ACATACGCCG	CTGGCTCAAC	GAGAACGGGC	CGCAGTTATG	CCCTATTATG	GCGCGGTGTA
AGCAGAACTT	TAAAAGTGCT	CATCATTGGA	AAACGTTCTT	CGGGGCGAAA	ACTCTCAAGG	ATCTTACCGC	TGTTGAGATC
TCGTCTTGAA	ATTTTCACGA	GTAGTAACCT	TTTGCAAGAA	GCCCCGCTTT	TGAGAGTTCC	TAGAATGGCG	ACAACTCTAG
CAGTTCGATG	TAACCCACTC	GTGCACCCAA	CTGATCTTCA	GCATCTTTTA	CTTTCACCAG	CGTTTCTGGG	TGAGCAAAA
GTCAAGCTAC	ATTGGGTGAG	CACGTGGGTT	GACTAGAAGT	CGTAGAAAAT	GAAAGTGGTC	GCAAAGACCC	ACTCGTTTTT
CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	TAAGGGCGAC	ACGGAAATGT	TGAATACTCA	TACTCTTCCT	TTTTCAATAT
GTCCTTCCGT	TTTACGGCGT	TTTTTCCCTT	ATTCCCGCTG	TGCCTTTACA	ACTTATGAGT	ATGAGAAGGA	AAAAGTTATA
TATTGAAGCA	TTTATCAGGG	TTATTGTCTC	ATGAGCGGAT	ACATATTTGA	ATGTATTTAG	AAAAATAAAC	AAATAGGGGT
ATAACTTCGT	AAATAGTCCC	AATAACAGAG	TACTCGCCTA	TGTATAAACT	TACATAAATC	TTTTTATTTG	TTTATCCCCA
TCCGCGCACA	TTTCCCCGAA	AAGTGCCACC	TGACGTCTAA	GAAACCATTA	TTATCATGAC	ATTAACCTAT	AAAAATAGGC
AGGCGCGTGT	AAAGGGGCTT	TTCACGGTGG	ACTGCAGATT	CTTTGGTAAT	AATAGTACTG	TAATTGGATA	TTTTTATCCG
	TAGAAGTGGA TTACCAATGC AATGGTTACG TGTAGATAAC ACATCTATTG CCAGATTTAT GGTCTAAATA GTCTATAAT GCATCGTGGT CGTAGCACCA CCATGGTTATG GGTACAACA CATGGTTATG GGTACAACA CATGGTTATG GGTACAACA CATGGTTATG GGTACAACA CAGGAACTT TCGTCTTGAA CAGGTCATT TCGTCTTGAA CAGGAACGCA CAGGAACGAAC	TAGAAGTGGA TCTAGGAAAA TTACCAATGC TTAATCAGTG AATGGTTACG AATTAGTCAC TGTAGATAAC TACGATACGG ACATCTATTG ATGCTATGCC CCAGATTTAT CAGCAATAAA GGTCTAAATA GTCGTTATTT GTCTATTAAT TGTTGCCGGG CGTAGCACCA CAGTGCGAGC CCCATGTTGT GCAAAAAAGC GGTACAACA CGTTTTTTCG CATGGTTATG GCAAAAAAGC GGTACAATAC CGTCGTGACG CCAAGTCATT CTGAGAATAG GGTTCAGTAA GACTCTTATC AGCAGAACTT TAAAAGTGCT TCGTCTTGAA ATTTCACGA CAGGTAGCAC ATTGGGTGAG CAGGAAGCA AAATGCCGCA GTCCTTCCGT TTTACGGCGT TATTGAAGCA ATTTGGGTGAG TATTGAAGCA TTTACCGCGAA AGGCGCGCACA TTTCCCCGAA AGGCGCGTGT AAAGGGGCTT GTATCACGAG GCCCTTTCGT GTATCACGAG GCCCTTTCGT	TAGAAGTGGA TCTAGGAAAA TTTAATTTT TTACCAATGC TTAATCAGTG AGGCACCTAT AATGGTTACG AATTAGTCAC TCCGTGGATA TGTAGATAAC TACGATACGG GAGGGCTTAC ACATCTATTG ATGCTATGCC CTCCCGAATG CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGTCTAAATA TGTTGCCGGG AAGCTAGAGT CAGATAATTA ACAACGGCCC TTCGATCTCA GCATCGTGGT GTCACGCTCG TCGTTTGGTA CGTAGCACCA CAGTGCGAGC AGCAAACCAT CCCATGTTGT GCAAAAAAGC GGTTAGCTCC GGGTACAACA CGTTTTTCG CCAATCGAGG CATGGTTATG GCAAAAAAGC GGTTAGCTCC GTACCAATAC CGTCGTGACG ATAATTCTCT GTACCAATAC CGTCGTGACG TATTAAGAGA CCAAGTCATT CTGAGAATAG TGTATGCGGC GGTTCAGTAA GACTCTTATC ACATACGCCG AGCAGAACTT TAAAAGTGCT CATCATTGGA TCGTCTTGAA ATTTTCACGA GTAGTAACCT CAGGAAGCTA TAACCCACTC GTGCACCCAA GTCAAGCTAC ATTGGGTGAG TTTTTCCCTT TATTGAAGCA TTTACCGCGA AAAAAGGGAA TCCGCGCACAA TTTTCCCCGAA AAGTGCCCCCACA TCCGCGCACAA TTTTCCCCGAA AAGTGCCCCCCACA TCCGCGCACAA TTTTCCCCGAA AAGTGCCCCCCACA TCCGCGCACAA TTTTCCCCGGAA AAGTGCCCCCCACA TCCGCGCACAA TTTTCCCCGGAA AAGTGCCCCCCACACACACACCCCCACACACACCCCCACACACAC	TAGAAGTGGA TCTAGGAAAA TTTAATTTT ACTTCAAAAT TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC ACATCTATTG ATGCTATGCC CTCCCGAATG GTAGACCGGG CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG GGTCTAAATA GTCGTTATTT GGTCGGTCGG CCTTCCCGGC GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CAGATAATTA ACAACGGCCC TTCGATCTCA TTCATCAAGC GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CGTAGCACCA CAGTGCGAGC AGCAAACCAT ACCGAAGTAA CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC GGGTACAACA CGTTTTTTCG CCAATCGAGG AAGCCAGGAG CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG GTACCAATAC CGTCGTGACG TATTAAGAGA ATGACAGTAC CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG GGTTCAGTAA GACTCTTATC ACATACGCCG CTGGCTCAAC AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT TCGTCTTGAA ATTTTCACGA GTAGTAACCT TTTGCAAGAA CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GTCAAGCTAC ATTGGGTGAG TTATTTCCCTT ATTGCAGAGA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC GTCCTTCCGT TTTACGGCT TTTTTCCCTT ATTGCAGAGA TATTGAAGCA TTTATCAGGG TTATTTCCCT ATGAGCGGAT ATAACTTCGT AAAATAGTCCC AAAAAAAGGGAA TAAGGGCGAC ATTACCGCTA TCCGCGCACA TTTCCCCGAA AAAAAGGGAA TAAGGGCGAC ATTACTTCGT AAAAGGGCTT TTTCACGGTG ACTGCAGATT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGACCTCTAA AGGCGGGTGT AAAGGGGCTT TTCACGGTGG ACTGCAGATT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGACCCTAA AGGCGCGTGT AAAGGGGCTT TTCACGGTGG ACTGCAGATT GTATCACGAG GCCCTTTCGT C	TAGCAATGC TTAATCAGT AGGCACCTAT CTCAGCGATC TGTCTATTTC AATGGTAGAT ACTGAGATAC TACGATACC TCCGTGGATA GAGTCGCTAG ACAGTCAAAAG GAGTCAGTAAAAG CAGATAAAAG CTCCCGAATG GTAGACCGGG GTCACGACGT CCCAGATTATT ATGCTATGCC CTCCCGAATG GTAGACCGGG GTCACGACGT CCAGATTATT CAGCAATAAA CCAGCCAGCC GGAAGGGCCC AGCGCAGAG GTCACGACGT CCCAGATTAT GGTCGTTATTT GGTCGGTCGG CCTTCCCGGC TCGCGTCTTC CAGTTAAATA GTCGTTATTT GGTCGGTCGG CCTTCCCGGC TCGCGTCTTC GTCATTAAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA CAGATAATTA ACAACGGCCC TTCGATCTCA TTCATCAAGC GGTCAATTAT CAGATAATTA ACAACGGCCC TCGGTTGGTA ACCGAAGTAA GTCGAGCCA CAGTGCGAGC AGCAAACCAT ACCGAAGTAA GTCGAGCCA CAGTGCGAGC AGCAAACCAT ACCGAAGTAA GTCGAGCCAC CAGTGCGAGC AGCAAACCAT ACCGAAGTAA GTCGAGCCAC CAGTGTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA CGTTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA CGTTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA CGTAGTATA CAGATACGCCG CAGCGAGTTC CCATCCGTAA GACCAATAC CGTCGTGACG TATTAAAGAGA ATGACAGTAC GGTAGCAATAC CGTCGTTACA CAATACGCCG CTGGCTCAAC GAGAAACGGGC ACCAAGTCAT TAAAAGTGCT CAATACGCCG CTGGCTCAAC GAGAAACGGGC ACCAGAACTT TAAAAGTGCC CAATCATGGA AAACGTTCTT CGGGGCGAAA TCGTCTTGAA ATTTCACAGA GTAGTAACCT TTTGCAAGAA GCCCCGCTTT CAGTCCGTAA ATTTCACGAC GTAGTAACCT TTTGCAAGAAA GCCCGCGTTT CAGTCAAGCTAC ATTTCAGGCGA AAAAAGGGAA TAAGCGCCG CTGGCTCAAC GAGAAACGGCC CAGTTCAGTAA ATTTCACGAC GTAGTAACCT TTTGCAAGCTA ATTTCACGACA TTTTCCCCTT ATTCCCCTT ATTCCCCTT TTTACGGCGT TTTTTCCCTT ATTCCCCTT TTTACAGAGA TTTTTCCCCTA AAAAAAACCATTAAACTTCGT AAAATAGTCCC AAAAAAAGGGAA TAAGCGCGAA ACATATTTGAAACTTCGT TAAAAGTCCC AAAAAAAAGGGAA TAAGCGCGAA ACATATTTGAAACTTCGT AAAATAGTCCC AAAAAAAAGGGAA TAACCGCCTA TGTATAAAACTACAGAC TTCACCCCTA AAAAAAAAAA	TAGRAGTGGA TCTAGGAAAA TTTAATTTT ACTTCAAAAT TTAGTTAGAT TCCATATATA TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG ACAGATAAAG CAAGTAGGTA TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC ACACTCTATTG ATGCTAATGCC CTCCCGAATG GTAGACCGGG GTCACGACCT TACTATGGCG CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTCCAGA GGTCTAAATA GTCGTTATTT GGTCGGTGGG CCTTCCCGGC TCGCGTCTTC ACCAGGACGT GTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCGA CCAGATAATTA ACAACGGCCC TTCGATCTCA TTCATCAAGG GGTCAATTAT CAAACGCGTT GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTCCATC GGTCAAATTAT CAAACGCGTT CCATGGTGGT GTCACGCTCG TCGTTTGGTA TGGCTCCATT CAGCTCCGGT TCCCAACGAT CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGCAGCACA CGGTACAACA CGTTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA GTCTTCATTC CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCGTAAG GTCTTCATTC CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCGTAA GATGCTTTC GTACCAATAC CGTCGTGACG TATTAAGAGA ATGACAGTAC GGTAGGAACA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTG GCTAGCAACA GTCTTCATTC GGTCAGGAACTT TAAAAGTGCT CATCATTGGA AAGCCAGTAC GGAGAACGGGC CGCAGTTATG CCAGGTAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCT CGGGGCGAAA ACTCTCAAGG CCAGGAACGTA TAAAAGTGCT CATCATTGGA AAACGTTCT CGGGGCGAAA ACTCTCAAGG CCAGGAAGCTA TAAAAGTGCT CATCATTGGA AAACGTTCT CGGGGCGAAA ACTCTCAAGG CCAGGAAGCTA AATTGGGTGAG CACGTGGGTT TTTGCAAGAT CGTCAGAAAAT GAAACTCCA CAGGAAGCAA AATGCCCAA AAAAAGGGAA TAAGCGCGC TTTTCACCCAG GTCATCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CATTACACAG GTCCTTCCGT TTTACGGCG TTTTTCCCCTT ATTGCAGAACT TTTGCAAAAT CAAAAGTGGTC TATTGAAGCA TTTACCAGAG TTATTTGCTC ATGACCGCTA TGTATAAAACT TACATAATACA TATTGAAGCA TTTACCAGAG TTATTTGCTC ATGACCGCTA TGTATAAAACT TACATAAAACT TATTGAAGCA TTTACCAGAG TTATTTGCTC ATGACCGCTA TGTATAAACT TACATAATATC TATTGAAGCA TTTACCAGAG TTATTTGCTC ATGACCGCTA TGTATATAACT TACATAAAACAAAAC	GGTTCAGTAA GACTCTTATC ACATACGCCG CTGGCTCAAC GAGAACGGCC CGCAGTTATG CCCTATTATG AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TCGTCTTGAA ATTTTCACGA GTAGTAACCT TTTGCAAGAAA GCCCCGCTTT TGAGAGTTCC TAGAATGGCG CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG GTCAAGCTAC ATTGGGTGAG CACGTGGGTT GACTAGAAGT CGTAGAAAAT GAAAGTGGTC GCAAAGACCC CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT GTCCTTCCGT TTTACGGCGT TTTTTCCCTT ATTCCCGCTG TGCCTTTACA ACTTATGAGT ATGAGAAGGA TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAAATAAACC ATAACTTCGT AAATAGTCCC AATAACAGAG TACTCGCCTA TGTATAAACT TACATAAATC TTTTTATTTG TCCGCGCCACA TTTCCCCGAA AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AGGCGCGTGT AAAGGGGCTT TTCACGGTGG ACTGCAGATT CTTTTGGTAAT AATAGTACTG TAATTGGATA GTATCACGAG GCCCTTTCGT C

FIGURE 6



pCMV-II

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT
	AGCGCGCAAA	GCCACTACTG	CCACTTTTGG	AGACTGTGTA	CGTCGAGGGC	CTCTGCCAGT	GTCGAACAGA	CATTCGCCTA
81	GCCGGGAGCA	GACAAGCCCG	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA
	CGGCCCTCGT	CTGTTCGGGC	AGTCCCGCGC	AGTCGCCCAC	AACCGCCCAC	AGCCCCGACC	GAATTGATAC	GCCGTAGTCT
161	GCAGATTGTA	CTGAGAGTGC	ACCATATGAA	GCTTTTTGCA	AAAGCCTAGG	CCTCCAAAAA	AGCCTCCTCA	CTACTTCTGG
	CGTCTAACAT	GACTCTCACG	TGGTATACTT	CGAAAAACGT	TTTCGGATCC	GGAGGTTTTT	TCGGAGGAGT	GATGAAGACC
241	AATAGCTCAG	AGGCCGAGGC	GGCCTCGGCC	TCTGCATAAA	TAAAAAAAT	TAGTCAGCCA	TGGGGCGGAG	AATGGGCGGA
	TTATCGAGTC	TCCGGCTCCG	CCGGAGCCGG	AGACGTATTT	ATTTTTTTA	ATCAGTCGGT	ACCCCGCCTC	TTACCCGCCT
321	ACTGGGCGGG	GAGGGAATTA	TTGGCTATTG	GCCATTGCAT	ACGTTGTATC	TATATCATAA	TATGTACATT	TATATTGGCT
	TGACCCGCCC	CTCCCTTAAT	AACCGATAAC	CGGTAACGTA	TGCAACATAG	ATATAGTATT	ATACATGTAA	ATATAACCGA
401	CATGTCCAAT	ATGACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT
	GTACAGGTTA	TACTGGCGGT	ACAACTGTAA	CTAATAACTG	ATCAATAATT	ATCATTAGTT	AATGCCCCAG	TAATCAAGTA
481	AGCCCATATA	TGGAGTTCCG	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	CCCGCCCATT
	TCGGGTATAT	ACCTCAAGGC	GCAATGTATT	GAATGCCATT	TACCGGGCGG	ACCGACTGGC	GGGTTGCTGG	GGGCGGGTAA
561	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT
	CTGCAGTTAT	TACTGCATAC	AAGGGTATCA	TTGCGGTTAT	CCCTGAAAGG	TAACTGCAGT	TACCCACCTC	ATAAATGCCA
641	AAACTGCCCA	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTCCGCCC	CCTATTGACG	TCAATGACGG	TAAATGGCCC
	TTTGACGGGT	GAACCGTCAT	GTAGTTCACA	TAGTATACGG	TTCAGGCGGG	GGATAACTGC	AGTTACTGCC	ATTTACCGGG
721	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	CGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC
	CGGACCGTAA	TACGGGTCAT	GTACTGGAAT	GCCCTGAAAG	GATGAACCGT	CATGTAGATG	CATAATCAGT	AGCGATAATG
801	CATGGTGATG	CGGTTTTGGC	AGTACACCAA	TGGGCGTGGA	TAGCGGTTTG	ACTCACGGGG	ATTTCCAAGT	CTCCACCCA
	GTACCACTAC	GCCAAAACCG	TCATGTGGTT	ACCCGCACCT	ATCGCCAAAC	TGAGTGCCC	TAAAGGTTCA	GAGGTGGGGT
881	TTGACGTCAA AACTGCAGTT	TGGGAGTTTG ACCCTCAAAC	TTTTGGCACC	AAAATCAACG	GGACTTTCCA CCTGAAAGGT	AAATGTCGTA TTTACAGCAT	ATAACCCCGC TATTGGGGCG	CCCGTTGACG GGGCAACTGC
961	CAAATGGGCG GTTTACCCGC	GTAGGCGTGT CATCCGCACA	ACGGTGGGAG TGCCACCCTC	GTCTATATAA CAGATATATI	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG GCAGTCTAGC	CCTGGAGACG GGACCTCTGC
1041	CCATCCACGO	TGTTTTGACC ACAAAACTGG	TCCATAGAAG AGGTATCTTC	ACACCGGGAC TGTGGCCCTG	CGATCCAGCC GCTAGGTCGG	TCCGCGGCCGGCCGG	GGAACGGTGC CCTTGCCACG	ATTGGAACGC TAACCTTGCG
1121	GGATTCCCCG CCTAAGGGGC	TGCCAAGAGT ACGGTTCTCA	GACGTAAGTA CTGCATTCAT	CCGCCTATAC GGCGGATATC	ACTCTATAGO TGAGATATCO	G CACACCCCTT	TGGCTCTTATA A ACCGAGAATA	GCATGCTATA CGTACGATAT
1201	CTGTTTTTGC	CTTGGGGCCT	TATACACCCCC	GCTCCTTATO CGAGGAATAO	CTATAGGTGA GATATCCACT	TGGTATAGCT	TAGCCTATAC	GTGTGGGTTA CACACCCAAT
1281	TTGACCATTA AACTGGTAA1	TTGACCACTO	CCCTATTGG:	GACGATACTI CTGCTATGA	TCCATTACTA A AGGTAATGAT	A ATCCATAAC	A TGGCTCTTTC T ACCGAGAAAC	CCACAACTAT GGTGTTGATA
1361	CTCTATTGGC GAGATAACCC	TATATGCCA	A TACTCTGTCG	TTCAGAGACT AAGTCTCTG	GACACGGACT CTGTGCCTG	CTGTATTTT	I ACAGGATGGO A TGTCCTACCO	G GTCCATTTAT CAGGTAAATA
1441	TATTTACAA/ ATAAATGTT	TTCACATATA AAGTGTATA	A CAACAACGC	GTCCCCCGT	GCCGCAGTT	TTATTAAAC A AATAATTTG	A TAGCGTGGGAT ATCGCACCC	A TCTCCGACAT C AGAGGCTGTA

pCMV-II

FIGURE 7 - Page 2

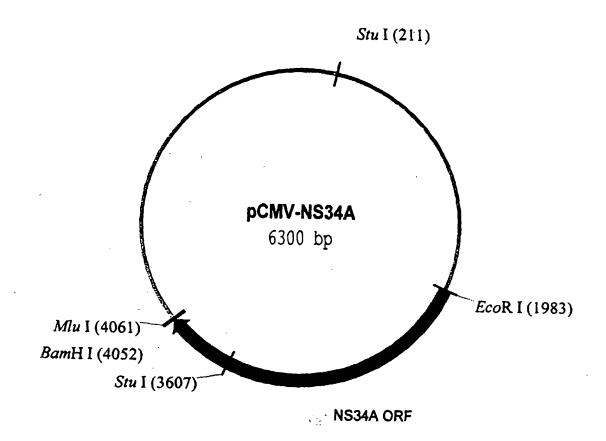
1521	CTCGGGTACG GAGCCCATGC	TGTTCCGGAC ACAAGGCCTG	ATGGGCTCTT TACCCGAGAA	CTCCGGTAGC GAGGCCATCG	GGCGGAGCTT CCGCCTCGAA	CCACATCCGA GGTGTAGGCT	GCCCTGGTCC CGGGACCAGG	CATCCGTCCA GTAGGCAGGT
1601	GCGGCTCATG CGCCGAGTAC	GTCGCTCGGC CAGCGAGCCG	AGCTCCTTGC TCGAGGAACG	TCCTAACAGT AGGATTGTCA	GGAGGCCAGA CCTCCGGTCT	CTTAGGCACA GAATCCGTGT	GCACAATGCC CGTGTTACGG	CACCACCACC GTGGTGGTGG
1681	AGTGTGCCGC TCACACGGCG	ACAAGGCCGT TGTTCCGGCA	GGCGGTAGGG CCGCCATCCC	TATGTGTCTG ATACACAGAC	AAAATGAGCT TTTTACTCGA	CGGAGATTGG GCCTCTAACC	GCTCGCACCT CGAGCGTGGA	GGACGCAGAT CCTGCGTCTA
1761	GGAAGACTTA CCTTCTGAAT	AGGCAGCGGC TCCGTCGCCG	AGAAGAAGAT TCTTCTTCTA	GCAGGCAGCT CGTCCGTCGA	GAGTTGTTGT CTCAACAACA	ATTCTGATAA TAAGACTATT	GAGTCAGAGG CTCAGTCTCC	TAACTCCCGT ATTGAGGGCA
1841	TGCGGTGCTG ACGCCACGAC	TTAACGGTGG AATTGCCACC	AGGGCAGTGT TCCCGTCACA	AGTCTGAGCA TCAGACTCGT	GTACTCGTTG CATGAGCAAC	CTGCCGCGCG GACGGCGCGC	CGCCACCAGA GCGGTGGTCT	CATAATAGCT GTATTATCGA
							EcoRI	
1921	GACAGACTAA CTGTCTGATT	CAGACTGTTC GTCTGACAAG	CTTTCCATGG GAAAGGTACC	GTCTTTTCTG CAGAAAAGAC	CAGTCACCGT GTCAGTGGCA	CGTCGACCTA GCAGCTGGAT	AGAATTCAGA	CTCGAGCAAG GAGCTCGTTC
	XbaI		Bami				•	
2001	TCTAGAAAGG AGATCTTTCC	CGCGCCAAGA GCGCGGTTCT	TATCAAGGAT ATAGTTCCTA	CCACTACGCG	TTAGAGCTCG	CTGATCAGCC GACTAGTCGG	TCGACTGTGC AGCTGACACG	CTTCTAGTTG GAAGATCAAC
2081	CCAGCCATCT GGTCGGTAGA	GTTGTTTGCC CAACAAACGG	CCTCCCCGT GGAGGGGGCA	GCCTTCCTTG CGGAAGGAAC	ACCCTGGAAG TGGGACCTTC	GTGCCACTCC CACGGTGAGG	CACTGTCCTT GTGACAGGAA	TCCTAATAAA AGGATTATTT
2161	ATGAGGAAAT TACTCCTTTA	TGCATCGCAT ACGTAGCGTA	TGTCTGAGTA ACAGACTCAT	GGTGTCATTC CCACAGTAAG	TATTCTGGGG ATAAGACCCC	GGTGGGGTGG CCACCCCACC	GGCAGGACAG CCGTCCTGTC	CAAGGGGGAG GTTCCCCCTC
2241	GATTGGGAAG CTAACCCTTC	ACAATAGCAG TGTTATCGTC	GCATGCTGGG CGTACGACCC	GAGCTCTTCC CTCGAGAAGG	GCTTCCTCGC CGAAGGAGCG	TCACTGACTC AGTGACTGAG	GCTGCGCTCG CGACGCGAGC	GTCGTTCGGC CAGCAAGCCG
2321	TGCGGCGAGC ACGCCGCTCG	GGTATCAGCT CCATAGTCGA	CACTCAAAGG GTGAGTTTCC	CGGTAATACG GCCATTATGC	GTTATCCACA CAATAGGTGT	GAATCAGGGG CTTAGTCCCC	ATAACGCAGG TATTGCGTCC	AAAGAACATG TTTCTTGTAC
2401	TGAGCAAAAG ACTCGTTTTC	GCCAGCAAAA CGGTCGTTTT	GGCCAGGAAC	CGTAAAAAGG GCATTTTTCC	CCGCGTTGCT GGCGCAACGA	GGCGTTTTTC	CATAGGCTCC GTATCCGAGG	GCCCCCTGA CGGGGGGACT
2481	CGAGCATCAC GCTCGTAGTG	AAAAATCGAC TTTTTAGCTG	GCTCAAGTCA CGAGTTCAGT	GAGGTGGCGA	AACCCGACAG TTGGGCTGTC	GACTATAAAG CTGATATTTC	ATACCAGGCG TATGGTCCGC	TTTCCCCCTG AAAGGGGGAC
2561	GAAGCTCCCT CTTCGAGGGA	CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGC1	TACCGGATAC	CTGTCCGCCT GACAGGCGGA	TTCTCCCTTC AAGAGGGAAG	GGGAAGCGTG CCCTTCGCAC
2641	GCGCTTTCTC CGCGAAAGAG	AATGCTCACG TTACGAGTGC	CTGTAGGTAT GACATCCATA	CTCAGTTCGC	TGTAGGTCGT ACATCCAGCA	TCGCTCCAAG AGCGAGGTTC	CTGGGCTGTG GACCCGACAC	TGCACGAACC ACGTGCTTGG
2721	CCCCGTTCAG GGGGCAAGTC	CCCGACCGCT GGGCTGGCGI	GCGCCTTATC	CGGTAACTA1	CGTCTTGAG1	CCAACCCGGT GGTTGGGCCA	AAGACACGAC	TTATCGCCAC AATAGCGGTG
2801	TGGCAGCAGC ACCGTCGTCG	CACTGGTAAC GTGACCATTC	AGGATTAGCA TCCTAATCGT	GAGCGAGGTA	A TGTAGGCGG1	GCTACAGAGT CGATGTCTCA	TCTTGAAGTO	G GTGGCCTAAC CACCGGATTG
2881	TACGGCTACA ATGCCGATGT	CTAGAAGGAC GATCTTCCTC	AGTATTTGGT	ATCTGCGCT(TGCTGAAGCG ACGACTTCGG	AGTTACCTTC TCAATGGAAC	GGAAAAAGAG CCTTTTTCTC	TTGGTAGCTC AACCATCGAG
_		·						

٠,٠:-

pCMV-II

TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT	TACGCGCAGA	AAAAAAGGAT
AACTAGGCCG	TTTGTTTGGT	GGCGACCATC	GCCACCAAAA	AAACAAACGT	TCGTCGTCTA	ATGCGCGTCT	TTTTTTCCTA
CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	TTTGGTCATG
GAGTTCTTCT	AGGAAACTAG	AAAAGATGCC	CCAGACTGCG	AGTCACCTTG	CTTTTGAGTG	CAATTCCCTA	AAACCAGTAC
AGATTATCAA	AAAGGATCTT	CACCTAGATC	CTTTTAAATT	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA
TCTAATAGTT	TTTCCTAGAA	GTGGATCTAG	GAAAATTTAA	TTTTTACTTC	AAAATTTAGT	TAGATTTCAT	ATATACTCAT
AACTTGGTCT	GACAGTTACC	AATGCTTAAT	CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT	ATTTCGTTCA	TCCATAGTTG
TTGAACCAGA	CTGTCAATGG	TTACGAATTA	GTCACTCCGT	GGATAGAGTC	GCTAGACAGA	TAAAGCAAGT	AGGTATCAAC
CCTGACTCCC	CGTCGTGTAG	ATAACTACGA	TACGGGAGGG	CTTACCATCT	GGCCCCAGTG	CTGCAATGAT	ACCGCGAGAC
GGACTGAGGG	GCAGCACATC	TATTGATGCT	ATGCCCTCCC	GAATGGTAGA	CCGGGGTCAC	GACGTTACTA	TGGCGCTCTG
CCACGCTCAC	CGGCTCCAGA	TTTATCAGCA	ATAAACCAGC	CAGCCGGAAG	GGCCGAGCGC	AGAAGTGGTC	CTGCAACTTT
GGTGCGAGTG	GCCGAGGTCT	AAATAGTCGT	TATTTGGTCG	GTCGGCCTTC	CCGGCTCGCG	TCTTCACCAG	GACGTTGAAA
ATCCGCCTCC	ATCCAGTCTA	TTAATTGTTG	CCGGGAAGCT	AGAGTAAGTA	GTTCGCCAGT	TAATAGTTTG	CGCAACGTTG
TAGGCGGAGG	TAGGTCAGAT	AATTAACAAC	GGCCCTTCGA	TCTCATTCAT	CAAGCGGTCA	ATTATCAAAC	GCGTTGCAAC
TTGCCATTGC	TACAGGCATC	GTGGTGTCAC	GCTCGTCGTT	TGGTATGGCT	TCATTCAGCT	CCGGTTCCCA	ACGATCAAGG
AACGGTAACG	ATGTCCGTAG	CACCACAGTG	CGAGCAGCAA	ACCATACCGA	AGTAAGTCGA	GGCCAAGGGT	TGCTAGTTCC
CGAGTTACAT	GATCCCCCAT	GTTGTGCAAA	AAAGCGGTTA	GCTCCTTCGG	TCCTCCGATC	GTTGTCAGAA	GTAAGTTGGC
GCTCAATGTA	CTAGGGGGTA	CAACACGTTT	TTTCGCCAAT	CGAGGAAGCC	AGGAGGCTAG	CAACAGTCTT	CATTCAACCG
CGCAGTGTTA	TCACTCATGG	TTATGGCAGC	ACTGCATAAT	TCTCTTACTG	TCATGCCATC	CGTAAGATGC	TTTTCTGTGA
GCGTCACAAT	AGTGAGTACC	AATACCGTCG	TGACGTATTA	AGAGAATGAC	AGTACGGTAG	GCATTCTACG	AAAAGACACT
CTGGTGAGTA	CTCAACCAAG	TCATTCTGAG	AATAGTGTAT	GCGGCGACCG	AGTTGCTCTT	GCCCGGCGTC	AATACGGGAT
GACCACTCAT	GAGTTGGTTC	AGTAAGACTC	TTATCACATA	CGCCGCTGGC	TCAACGAGAA	CGGGCCGCAG	TTATGCCCTA
AATACCGCGC	CACATAGCAG	AACTTTAAAA	GTGCTCATCA	TTGGAAAACG	TTCTTCGGGG	CGAAAACTCT	CAAGGATCTT
TTATGGCGCG	GTGTATCGTC	TTGAAATTTT	CACGAGTAGT	AACCTTTTGC	AAGAAGCCCC	GCTTTTGAGA	GTTCCTAGAA
ACCGCTGTTG	AGATCCAGTT	CGATGTAACC	CACTCGTGCA	CCCAACTGAT	CTTCAGCATC	TTTTACTTTC	ACCAGCGTTT
TGGCGACAAC	TCTAGGTCAA	GCTACATTGG	GTGAGCACGT	GGGTTGACTA	GAAGTCGTAG	AAAATGAAAG	TGGTCGCAAA
CTGGGTGAGC GACCCACTCG	AAAAACAGGA TTTTTGTCCT	AGGCAAAATG	CCGCAAAAA GGCGTTTTT	GGGAATAAGG	GCGACACGGA CGCTGTGCCT	AATGTTGAAT TTACAACTTA	ACTCATACTC TGAGTATGAG
TTCCTTTTTC AAGGAAAAAG	AATATTATTG	AAGCATTTAT	CAGGGTTATT	GTCTCATGAG CAGAGTACTC	CGGATACATA CCCTATGTAT	TTTGAATGTA AAACTTACAT	TTTAGAAAAA AAATCTTTTT
TAAACAAATA ATTTGTTTAT	GGGGTTCCGC	GCACATTTCC CGTGTAAAGC	CCGAAAAGTG GGCTTTTCAC	CCACCTGACG GGTGGACTGC	TCTAAGAAAC AGATTCTTTG	CATTATTATO	ATGACATTAA TACTGTAATT
CCTATAAAAA GGATATTTT	TAGGCGTATO	ACGAGGCCC1	TTCGTC AAGCAG				
	AACTAGGCCG CTCAAGAAGA GAGTTCTTCT AGATTATCAA TCTAATAGTT TTGAACCAGA CCTGACTCCC GGACTGAGGG CCACGCTCAC GGTGCGAGGG TTGCCATTGC AACGGTAACG CGAGTTACAT CCGCAGTCAC CGAGTTACAT GCTCAATGTA CTGGTGAGTA CTGGTTTG TTCCTTTTTC AAGGAAAAAA TTAAACAAATA ATTTGTTTAT CCTATAAAAA	AACTAGGCCG TTTGTTTGGT CTCAAGAAGA TCCTTTGATC GAGTTCTTCT AGGAAACTAG AGATTATCAA AAAGGATCTT TCTAATAGTT TTTCCTAGAA AACTTGGTCT GACAGTTACC TTGAACCAGA CTGTCAATGG CCTGACTCCC CGTCGTGTAG GGACTGAGGG GCAGCACATC CCACGCTCAC CGGCTCCAGA GGTGCGAGTG GCCGAGGTCT ATCCGCCTCC ATCCAGTCTA TAGGCGGAGG TAGGTCAGAT CGAGTTACAT GATCCCCAT GCTCAATGAT CTAGGCGTAC CGAGTTACAT GATCCCCAT GCTCAATGT AGTGAGTACC CTGGTGAGTA CTCAACCAAG GACCACTCAT GAGTTGGTTC AATACCGCGC CACATAGCAG TTATGGCGGC GTGTATCGTC ACCGCTGTTG AGATCCAGTT TGGCGACAAC TCTAGGTCAA CTGGGTGAGC AAAAACAGGA GACCCACTCG TTTTTGTCCT TTCCTTTTTC AATATTATTG AAGGAAAAAA TAGGCGTATC CCCTATAAAAAA TAGGCGTATC	AACTAGGCCG TTTGTTTGGT GGCGACCATC CTCAAGAAGA TCCTTTGATC TTTTCTACGG GAGTTCTTCT AGGAAACTAG AAAAGATGCC AGATTATCAA AAAGGATCTT CACCTAGATC TCTAATAGTT TTTCCTAGAA GTGGATCTAG AACTTGGTCT GACAGTTACC AATGCTTAAT TTGAACCAGA CTGTCAATGG TTACGAATTA CCTGACTCCC CGTCGTGTAG ATAACTACGA GGACTGAGGG GCAGCACATC TATTGATGCT ATCCGCCTCC ATCCAGA TTTATCAGCA GGTGCGAGG TAGGTCAGAT AAATAACCGT ATCCGCCTCC ATCCAGGTCTA TTAATTGTTG TAGGCGGAGG TAGGTCAGAT AATTAACAAC CGAGGTTACAT GATCCCCCAT GTGTGCAAA GCTCAATGT TACAGGCATC GTGTGTCACA ACGGTTACAT GATCCCCCAT GTTGTGCAAA GCTCAATGTA CTAAGGGGTA CAACACGTTT CGCAGTGTA TCACTCATGG TTATGGCAGC GCGTCACAAT AGTGAGTACC AATACCGTCG CTGGTGAGTA CTCAACCAAG TCATTCTGAG GACCACTCAT GAGTTGGTTC AGTAAGACTC AATACCGCCG CACATAGCAG TCATTCTGAG GACCACTCAT GAGTTCGTC TTGAAATTTT ACCGCTGTTG AGATCCAGTT CGATGTAACC CTGGGTGAGC AAAAACAGGA AGCTATATAAA TTATGGCGCG TTTTTTGTCCT TCCGTTTTAC CTGGGTGAGC AAAAACAGGA AGGCAAAATG GACCCACTCG TTTTTTGTCCT TCCGTTTTAC TTCCTTTTC AATATTATTG AAGCAATAACC CTGGTGAGC AAAAACAGGA AGGCAAAATG GACCCACTCG TTTTTTGTCCT TCCGTTTTAC TTCCTTTTTC AATATTATTG AAGCATTTAT AAGGAAAAAA TAGGCGTATC ACGAGGCCCCT CCTATAAAAAA TAGGCGTATC ACGAGGCCCCT CCTATAAAAAA TAGGCGTATC ACGAGGCCCCT	AACTAGGCCG TTTGTTTGGT GGCGACCATC GCCACCAAAA CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC GAGTTCTTCT AGGAAACTAG AAAAGATGCC CCAGACTGCG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT TCTAATAGTT TTTCCTAGAA GTGGATCTAG GAAAATTTAA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGGGAGGCA TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG GGACTGAGGG GCAGCACATC TATTGATGCT ATGCCCTCCC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC GGTGCGAGTG GCCGAGGTCT AAATAGTCGT TATTTGGTCG ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT TAGGCGGAGG TAGGTCAGAT AATTAACAAC GGCCCTTCGA TTGCCATTGC TACAGGCATC GTGGTGTCAC GCTCGTCGTT AACGGTAACG ATGTCCGTAG CACCACAGTG CGAGCAGCAA CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCAATGTA CTAGGGGGTA CAACACGTTT TTTCGCCAAT CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT GCGCTCACAAT AGTGAGTACC AATACCGTCG TGACGTATTA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATACCGTCAT AATACCGCCG CACATAGCAG AACTTCTGAG AATACCGTCAT AATACCGCGC CACATAGCAG TCATTCTGAG AATACCGTAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTATGGCGCG GTGTATCCTC TTGAAATTTT CACGAGTAGT TTATGGCGCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT TCCTTTTTC AAAAACAGGA AGGCAAAATG CCGCAAAAAAA GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT AAGGAAAAAAAAAAAAAAAAAAAAAAAA	AACTAGGCCG TTTGTTTGGT GGCGACCATC GCCACCAAAA AAACAAACGT CTCAAGAAGA TCCTTTGATC TTTCTACGG GGTCTGACGC TCAGTGGAAC GAGTTCTTCT AGGAAACTAG AAAAGATCC CCAGACTGCG AGTCACCTTG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAC TCTAATAGTT TTTCCTAGAA GTGGATCTAG GAAAATTTAA TTTTTACTTC AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCG GGATAGAGTC CCTGACTCCC CGTCGTGTAG ATACTACGA TACGGGAGGG CTTACCATCT GGACTGAGGG GCAGCACATC TATTGATGCT ATGCCCTCCC GAATGGTAGA CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT TATTTGGTCG GTCGGCCTTC ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGGT AGAGTAAGTA TAGGCGGAGG TAGGTCAGAT TAATTACTAG GCCCTTCGA TCTCATTCAT TTGCCATTGC TACAGGCATC GTGGTGTCAC GCTCGTCTT TGGTATGCT AACGGTAACG ATGTCCCGTAG CACCACAGTG CGAGCAGCAA ACCATACCGA CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG GCTCAATGTA CTAGGGGGTA CACCACAGTG CGAGCAGCAA ACCATACCGA CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA CGAGGAAGCC CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG GCTCAATGTA TACACCAAG TCATTCGAGA AACACGTTT TTTCGCCAAT CGAGGAAGCC CCGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCCTTTACTG GCCCACCACATGT GAGTTGGTC AATACCGTCA TTATCACATA CGCGCGCACCG GACCACTCAT GAGTTGGTTC AGTAAGACTC TTATCACATA CGCGCGCACGG AATACCGCGC CACATAGCAG AACTTTAAAA TTATGGCCGC GACATAGCAG AACTTTAAAAA AAACCGGCG GTGTATCGTC TTGAAATTTT CACCGATAAT TCCTTATCG ACCGCTTGTTG AGATCCAGTT CGATGTAACC CACTCGTGC CCCAACTGAT TTGGCCACACAC TCTAGGTCAA GCTACATTTA CACGAGTAAT TCCCTTATTCC TTCCTTTTC AACACAAG AGCCAATTTAAAA GTGCCAAAAAA GGGAATAACC TTATGGCCGC CACATAGCAG ACCTTTTACACACAA GCCCACTGAT TTGGCCACACC TCTTTGTCCT TCGGTTTTAC CCCCAACAAAAA GGGAATAACC TTATCGCCACCAC TCTAGGTCAA CTCACTTTAC CCCCCTTATTCC TTCCTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GCCCAACTGAT TAAACAAAAA GGGAATAAAC TCCGAAAATA CAGAGTAACT TAAACAAAAA GGGATACCA CCCAATTTCC CCCAAAAAAA CAGAGTACTC TAAACAAAAA GGGATACCA CCCACATTCC CCCAAAAAAA CAGAGTACTC TAAACAAAAA GGGGTTCCGC GCACATTTCC CCCAAAAAAT CAGAGTACTC TAAACAAAAA TAGGCGTATC ACGAGGCCCT TCCGCC CCTAATA	AACTAGGCCG TITGTTTGGT GGCGACCATC GCCACCAAAA AAACAAACGT TCGTCGTCTA CTCAAGAAGA TCCTTTGATC TITTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTAC GAGTTCTTCT AGGAAACTAG AAAAGATGCC CCAGACTGCG AGTCACCTG CTTTTGAGTG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA TCTAATAGTT TTTCCTAGAA GTGGATCTAG GAAAATTTAA TTTTTACTTC AAAATTTAGT AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG GGACTGAGGG GCAGCACATC TATTGATGCT ATGCCCTCCC GAAAAGGGC GCTACCATG GGACTGAGGG GCAGCACATC TATTGATGCT ATGCCCTCCC GAAAGGGC GCCGGAAG GGCCGAGGTC AAATAGTCGT TATTGGTCG GTCGGCCTTC CCGGGTCAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GCCCGAGGG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT TATTTGGTCG GTCGGCCTTC CCGGCTCCGG ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAGGCGGAGG TAGGTCAGAT AAATAACAAC GGCCCTTCGA TCTCATTCAT CAAGCGGTCA ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAGGCGGAGG TAGGTCAGAT AAATAACAAC GGCCCTTCGA TCTCATTCAT CAAGCGGTCA TTGCCATTGC TACAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT AACGGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCATCCGA AGTAAGTCA CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAACCGGTTA GCTCCTTCGG TCCTCCGATC GCTCAATGTA CTAGGGGGTA CAACACGGTT TTTCGCCAAT CGAGGAAGCC AGGAGGCTAG CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCCTTTACT TCATGCCATC GCCGTCACAAT AGTGGAGTACC AATACCGTCG TGACGTATAT AGAGAAACCA AGTACCGTA AATACCGCCC CACATAGCAG ACTTTAAAA AAACACTCAT GAGTTGGTTC AGTAAGACCC TTATCACATA AGGAAATGAC AGTACCGTAG AATACCGCCC CACATAGCAG ACTTTAAAA GTGCCATCA TTGGAAAACG TCTATCGCGAGA ACCACTCAT GAGTTGGTTC TGGATTAACC CCGCGTGTGA AAAAACAGGA ACCTTAAAAA AGGAATAGA CCCACTCGT TCAACCAAGGAACC TTATCACATC GAGTTACCTC TTGAAATTTT CACGCTAGCAG GGGAAACCCACTCG TTTTTTTCCCAATTAAAAA GGGAAAAAA GGGAAAAAC GCTAATATATTG GCCGCGTGCA CCCAACTGAT CTTACACATA CACGAGTAGT CTCAAGCAGA AATACCGCCC CACATAGCAG AACTTTAAAAA GTGCCAACTGAT CTCAAGAGACCC TCCGCTGTTGC AAAAACAGGA AGCCAATTTCC CCGCAAAAAAA GGGAAAAAG GCCACACTGA CTCAAGAGACC CACCTATGTAT CCCCAATAAA CACAATAAA CACAATAAC CACAATAAC CACCAATAAA CACAATAAA	TGGCGACAAC TCTAGGTCAA GCTACATTGG GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT CCCTTATTCC CGCTGTGCCT TTACAACTTA TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA AAGGAAAAAG TTATAATAAC TTCGTAAATA GTCCCAATAA CAGAGTACTC GCCTATGTAT AAACTTACAT TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC ATTTGTTTAT CCCCAAGGCG CGTGTAAAGG GGCTTTTCAC GGTGGACTGC AGATTCTTTG GTAATAATAG

FIGURE 8



pCMV-NS34A

	TCGCGCGTTT AGCGCGCAAA	CGGTGATGAC GCCACTACTG	GGTGAAAACC CCACTTTTGG	TCTGACACAT AGACTGTGTA	GCAGCTCCCG CGTCGAGGGC	
51	GAGACGGTCA CTCTGCCAGT	CAGCTIGTCT GTCGAACAGA	GTAAGCGGAT CATTCGCCTA	GCCGGGAGCA CGGCCCTCGT	GACAAGCCCG CTGTTCGGGC	
101	TCAGGGCGCG AGTCCCGCGC	TCAGCGGGTG AGTCGCCCAC	TTGGCGGGTG AACCGCCCAC	TCGGGGCTGG AGCCCCGACC	CTTAACTATG GAATTGATAC	
151	CGGCATCAGA GCCGTAGTCT	GCAGATTGTA CGTCTAACAT	CTGAGAGTGC GACTCTCACG	ACCATATGAA TGGTATACTT	GCTTTTTGCA CGAAAAACGT	
	St	uI				
201	AAAGCCTAGG	CCTCCAAAAA GGAGGTTTTT	AGCCTCCTCA TCGGAGGAGT	CTACTTCTGG GATGAAGACC	AATAGCTCAG TTATCGAGTC	
251	AGGCCGAGGC TCCGGCTCCG	GGCCTCGGCC CCGGAGCCGG	TCTGCATAAA AGACGTATTT	TAAAAAAAT ATTTTTTTA	TAGTCAGCCA ATCAGTCGGT	
301	TGGGGCGGAG ACCCCGCCTC	AATGGGCGGA TTACCCGCCT	ACTGGGCGGG TGACCCGCCC	GAGGGAATTA CTCCCTTAAT	TTGGCTATTG AACCGATAAC	
351	GCCATTGCAT CGGTAACGTA	ACGTTGTATC TGCAACATAG	TATATCATAA ATATAGTATT	TATGTACATT ATACATGTAA	TATATTGGCT ATATAACCGA	
401	CATGTCCAAT GTACAGGTTA	ATGACCGCCA TACTGGCGGT	TGTTGACATT ACAACTGTAA	GATTATTGAC CTAATAACTG	TAGTTATTAA ATCAATAATT	
451	TAGTAATCAA ATCATTAGTT	TTACGGGGTC AATGCCCCAG	ATTAGTTCAT TAATCAAGTA	AGCCCATATA TCGGGTATAT	TGGAGTTCCG ACCTCAAGGC	
501	CGTTACATAA GCAATGTATT	CTTACGGTAA GAATGCCATT	ATGGCCCGCC TACCGGGCGG	TGGCTGACCG ACCGACTGGC	CCCAACGACC GGGTTGCTGG	
551	CCCGCCCATT GGGCGGGTAA	GACGTCAATA CTGCAGTTAT	ATGACGTATG TACTGCATAC	TTCCCATAGT AAGGGTATCA	AACGCCAATA TTGCGGTTAT	
601	GGGACTTTCC CCCTGAAAGG	ATTGACGTCA TAACTGCAGT	ATGGGTGGAG TACCCACCTC	TATTTACGGT ATAAATGCCA	AAACTGCCCA TTTGACGGGT	
651	CTTGGCAGTA GAACCGTCAT	CATCAAGTGT GTAGTTCACA	ATCATATGCC TAGTATACGG	AAGTCCGCCC TTCAGGCGGG	CCTATTGACG GGATAACTGC	
701	TCAATGACGG AGTTACTGCC	TAAATGGCCC ATTTACCGGG	GCCTGGCATT CGGACCGTAA	ATGCCCAGTA TACGGGTCAT	CATGACCTTA GTACTGGAAT	
751	CGGGACTTTC GCCCTGAAAG	CTACTTGGCA GATGAACCGT	GTACATCTAC CATGTAGATG	GTATTAGTCA CATAATCAGT	TCGCTATTAC AGCGATAATG	
801	CATGGTGATG GTACCACTAC	CGGTTTTGGC GCCAAAACCG	AGTACACCAA TCATGTGGTT	TGGGCGTGGA ACCCGCACCT	TAGCGGTTTG ATCGCCAAAC	
851	ACTCACGGGG TGAGTGCCCC	ATTTCCAAGT TAAAGGTTCA	CTCCACCCCA GAGGTGGGGT	TTGACGTCAA AACTGCAGTT	TGGGAGTTTG ACCCTCAAAC	

pCMV-NS34A

901	TTTTGGCACC AAAACCGTGG	AAAATCAACG TTTTAGTTGC	GGACTTTCCA CCTGAAAGGT	AAATGTCGTA TTTACAGCAT	ATAACCCCGC TATTGGGGCG	
951	CCCGTTGACG GGGCAACTGC	CAAATGGGCG GTTTACCCGC	GTAGGCGTGT CATCCGCACA	ACGGTGGGAG TGCCACCCTC	GTCTATATAA CAGATATATT	
1001	GCAGAGCTCG CGTCTCGAGC	TTTAGTGAAC AAATCACTTG	CGTCAGATCG GCAGTCTAGC	CCTGGAGACG GGACCTCTGC	CCATCCACGC GGTAGGTGCG	
1051	TGTTTTGACC ACAAAACTGG	TCCATAGAAG AGGTATCTTC	ACACCGGGAC TGTGGCCCTG	CGATCCAGCC GCTAGGTCGG	TCCGCGGCCG AGGCGCCGGC	
1101	GGAACGGTGC CCTTGCCACG	ATTGGAACGC TAACCTTGCG	GGATTCCCCG CCTAAGGGGC	TGCCAAGAGT ACGGTTCTCA	GACGTAAGTA CTGCATTCAT	
1151	CCGCCTATAG GGCGGATATC	ACTCTATAGG TGAGATATCC	CACACCCCTT GTGTGGGGAA	TGGCTCTTAT ACCGAGAATA	GCATGCTATA CGTACGATAT	
1201	CTGTTTTTGG GACAAAAACC	CTTGGGGCCT GAACCCCGGA	ATACACCCCC TATGTGGGGG	GCTCCTTATG CGAGGAATAC	CTATAGGTGA GATATCCACT	
1251	TGGTATAGCT ACCATATCGA	TAGCCTATAG ATCGGATATC	GTGTGGGTTA CACACCCAAT	TTGACCATTA AACTGGTAAT	TTGACCACTC AACTGGTGAG	
1301	CCCTATTGGT GGGATAACCA	GACGATACTT CTGCTATGAA	TCCATTACTA AGGTAATGAT	ATCCATAACA TAGGTATTGT	TGGCTCTTTG ACCGAGAAAC	
1351	CCACAACTAT GGTGTTGATA	CTCTATTGGC GAGATAACCG	TATATGCCAA ATATACGGTT	TACTCTGTCC ATGAGACAGG	TTCAGAGACT AAGTCTCTGA	
1401	GACACGGACT CTGTGCCTGA	CTGTATTTT GACATAAAAA	ACAGGATGGG TGTCCTACCC	GTCCATTTAT CAGGTAAATA	TATTTACAAA ATAAATGTTT	
1451	TTCACATATA AAGTGTATAT	CAACAACGCC GTTGTTGCGG	GTCCCCGTG CAGGGGGCAC	CCCGCAGTTT GGGCGTCAAA	TTATTAAACA AATAATTTGT	
1501	TAGCGTGGGA ATCGCACCCT	TCTCCGACAT AGAGGCTGTA	CTCGGGTACG GAGCCCATGC	TGTTCCGGAC ACAAGGCCTG	ATGGGCTCTT TACCCGAGAA	. *
1551	CTCCGGTAGC GAGGCCATCG	GGCGGAGCTT CCGCCTCGAA	CCACATCCGA GGTGTAGGCT	GCCCTGGTCC CGGGACCAGG	CATCCGTCCA GTAGGCAGGT	
1601	GCGGCTCATG CGCCGAGTAC	GTCGCTCGGC CAGCGAGCCG	AGCTCCTTGC TCGAGGAACG	TCCTAACAGT AGGATTGTCA	GGAGGCCAGA CCTCCGGTCT	
1651	CTTAGGCACA GAATCCGTGT	GCACAATGCC CGTGTTACGG	CACCACCACC GTGGTGGTGG	AGTGTGCCGC TCACACGGCG	ACAAGGCCGT TGTTCCGGCA	
1701	GGCGGTAGGG CCGCCATCCC	TATGTGTCTG ATACACAGAC	AAAATGAGCT TTTTACTCGA	CGGAGATTGG GCCTCTAACC	GCTCGCACCT CGAGCGTGGA	
1751	GGACGCAGAT CCTGCGTCTA	GGAAGACTTA CCTTCTGAAT	AGGCAGCGGC TCCGTCGCCG	AGAAGAAGAT TCTTCTTCTA	GCAGGCAGCT CGTCCGTCGA	
1801	GAGTTGTTGT CTCAACAACA	ATTCTGATAA TAAGACTATT	GAGTCAGAGG CTCAGTCTCC	TAACTCCCGT ATTGAGGGCA	TGCGGTGCTG ACGCCACGAC	

pCMV-NS34A

1851		AGGGCAGTGT TCCCGTCACA						
1901		CATAATAGCT GTATTATCGA						
+2				EcoRI	M A	P		
1951		CAGTCACCGT GTCAGTGGCA						
	TCACGGCGTA	A Q Q CGCCCAGCAG GCGGGTCGTC	ACAAGGGGCC		CATAA		 	
		R D K GCCGGGACAA CGGCCCTGTT		GAGGGTGAGG	TCCAG	ATTGT		
_		A Q T E GCCCAAACCT CGGGTTTGGA		GTGCATCAAT		STGCT		
_		H G A CCACGGGGCC GGTGCCCCGG	GGAACGAGGA		ACCCA			
		M Y T AGATGTATAC TCTACATATG		CAAGACCTTG	TGGGC			
		G T R S GGTACCCGCT CCATGGGCGA	CATTGACACC			CTCGG		
		V T R GGTCACGAGG CCAGTGCTCC	CACGCCGATG		GCGCCG			
		G S L GGGGCAGCCT CCCCGTCGGA			CCTAC			
+2 2401	AGGCTCCTCG	G G P I GGGGGTCCGC CCCCCAGGCG	TGTTGTGCCC	CGCGGGGCAC		GGCA		
	TATTTAGGGC	A V C CGCGGTGTGC GCGCCACACG	ACCCGTGGAG	TGGCTAAGGC	GGTGG	ACTTT		
		E N L E AGAACCTAGA TCTTGGATCT		AGGTCCCCGG	TGTTC	ACGGA		

pcmv-ns34A FIGURE 9 - Page 4

+2 2551	N S S TAACTCCTCT ATTGAGGAGA	P P V V CCACCAGTAG GGTGGTCATC	TGCCCCAGAG	CTTCCAGGTG	GCTCACCTCC		
+2 2601 	H A P T ATGCTCCCAC TACGAGGGTG	G S G AGGCAGCGGC TCCGTCGCCG	AAAAGCACCA	AGGTCCCGGC	TGCATATGCA		
+2 2651	A Q G) GCTCAGGGCT CGAGTCCCGA	Y K V L ATAAGGTGCT TATTCCACGA	AGTACTCAAC	CCCTCTGTTG	CTGCAACACT		
2701 	G F G GGGCTTTGGT CCCGAAACCA	A Y M S GCTTACATGT CGAATGTACA	CCAAGGCTCA	TGGGATCGAT	CCTAACATCA		
2751	R T. G V GGACCGGGGT CCTGGCCCCA	GAGAACAATT	ACCACTGGCA	GCCCCATCAC	GTACTCCACC		
7 D U I	Y G K E TACGGCAAGT ATGCCGTTCA	TCCTTGCCGA	CGGCGGGTGC	TCGGGGGGGG	CTTATGACAT		
2851	I I C AATAATTTGT TTATTAAACA	GACGAGTGCC	ACTCCACGGA	TGCCACATCC	ATCTTGGGCA		-
2901	I G T V TTGGCACTGT AACCGTGACA	CCTTGACCAA	GCAGAGACTG	CGGGGGGGGAG	ACTGGTTGTG		
+2 951	L A T A CTCGCCACCG GAGCGGTGGC	CCACCCCTCC	GGGCTCCGTC	ACTGTGCCCC	ATCCCAACAT		
+2 3001	E E V CGAGGAGGTT GCTCCTCCAA	GCTCTGTCCA	CCACCGGAGA	GATCCCTTTT	TACGGCAAGG		, ,
1051	A I P L CTATCCCCCT GATAGGGGGA	CGAAGTAATC .	AAGGGGGGA	GACATCTCAT	CTTCTGTCAT		
101	S K K K TCAAAGAAGA AGTTTCTTCT	AGTGCGACGA .	ACTCGCCGCA	AAGCTGGTCG	CATTGGGCAT		
+2 151	N A V CAATGCCGTG GTTACGGCAC	GCCTACTACC (GCGGTCTTGA	CGTGTCCGTC	ATCCCGACCA		
201	S G D V GCGGCGATGT CGCCGCTACA	TGTCGTCGTG (GCAACCGATG	CCCTCATGAC	CGGCTATACC		

pCMV-NS34A

FIGURE 9 - Page 5

+2	G D F D S V I D C N T C V T Q T V	
	GGCGACTTCG ACTCGGTGAT AGACTGCAAT ACGTGTGTCA CCCAGACAGT CCGCTGAAGC TGAGCCACTA TCTGACGTTA TGCACACAGT GGGTCTGTCA	
_	D F S L D P T F T I E T I T L P CGATTTCAGC CTTGACCCTA CCTTCACCAT TGAGACAATC ACGCTCCCCC GCTAAAGTCG GAACTGGGAT GGAAGTGGTA ACTCTGTTAG TGCGAGGGGG	·
	Q D A V S R T Q R R G R T G R G K AAGATGCTGT CTCCCGCACT CAACGTCGGG GCAGGACTGG CAGGGGGAAG TTCTACGACA GAGGGCGTGA GTTGCAGCCC CGTCCTGACC GTCCCCCTTC	
	P G I Y R F V A P G E R P S G M F CCAGGCATCT ACAGATTGT GGCACCGGGG GAGCGCCCCT CCGGCATGTT GGTCCGTAGA TGTCTAAACA CCGTGGCCCC CTCGCGGGGA GGCCGTACAA	
	D S S V L C E C Y D A G C A W Y CGACTCGTCC GTCCTCTGTG AGTGCTATGA CGCAGGCTGT GCTTGGTATG GCTGAGCAGG CAGGAGACAC TCACGATACT GCGTCCGACA CGAACCATAC	
	E L T P A E T T V R L R A Y M N T AGCTCACGCC CGCCGAGACT ACAGTTAGGC TACGAGCGTA CATGAACACC TCGAGTGCGG GCGGCTCTGA TGTCAATCCG ATGCTCGCAT GTACTTGTGG	
3551	P G L P V C Q D H L E F W E G V F CCGGGGCTTC CCGTGTGCCA GGACCATCTT GAATTTTGGG AGGGCGTCTT GGCCCCGAAG GGCACACGGT CCTGGTAGAA CTTAAAACCC TCCCGCAGAA	
· +2	TGLTHID AHFLSQTKQ StuI	
3601	TACAGGCCTC ACTCATATAG ATGCCCACTT TCTATCCCAG ACAAAGCAGA ATGTCCCGGAG TGAGTATATC TACGGGTGAA AGATAGGGTC TGTTTCGTCT	
3651 	S G E N L P Y L V A Y Q A T V C A GTGGGGAGAA CCTTCCTTAC CTGGTAGCGT ACCAAGCCAC CGTGTGCGCT CACCCCTCTT GGAAGGAATG GACCATCGCA TGGTTCGGTG GCACACGCGA	
3701	R A Q A P P P S W D Q M W K C L I AGGGCTCAAG CCCCTCCCC ATCGTGGGAC CAGATGTGGA AGTGTTTGAT TCCCGAGTTC GGGGAGGGGG TAGCACCCTG GTCTACACCT TCACAAACTA	1.
	R L K P T L H G P T P L L Y R L TCGCCTCAAG CCCACCCTCC ATGGGCCAAC ACCCCTGCTA TACAGACTGG AGCGGAGTTC GGGTGGGAGG TACCCGGTTG TGGGGACGAT ATGTCTGACC	
3801	G A V Q N E I T L T H P V T K Y I GCGCTGTTCA GAATGAAATC ACCCTGACGC ACCCAGTCAC CAAATACATC CGCGACAAGT CTTACTTTAG TGGGACTGCG TGGGTCAGTG GTTTATGTAG	
	M T C M S A D L E V V T S T W V L ATGACATGCA TGTCGGCCGA CCTGGAGGTC GTCACGAGCA CCTGGGTGCT TACTGTACGT ACAGCCGGCT GGACCTCCAG CAGTGCTCGT GGACCCACGA	
3901	V G G V L A A L A A Y C L S T G CGTTGGCGGC GTCCTGGCTG CTTTGGCCGC GTATTGCCTG TCAACAGGCT GCAACCGCCC CATAACGGAC AGTTGTCCGA	

. : }_

pCMV-NS34A

T V G R CAT AGTGGGCAGG TA TCACCCGTCC R E V L Y GGG AAGTCCTCTA CCC TTCAGGAGAT	CAGCAGAACA R E F CCGAGAGTTC	CCGGGAAGCC GGCCCTTCGG D E M E	GGCAATCATA CCGTTAGTAT	
GGG AAGTCCTCTA	CCGAGAGTTC			
	GGCTCTCAAG		AAGAGTGCTA	
. MluI				
_	-			
	CTA CGCGTTAGAG GAT GCGCAATCTC GCC ATCTGTTGTT CGG TAGACAACAA CCA CTCCCACTGT GAG GAGGGTGACA CTG AGTAGGTGTC GAC CCTCCTAACC CCT CCTAACC CCT CCGTCCC CCT CCTACC CCT CCGCCCGC CCT CCGCCCC CCT CCGCCCC CCT CCGCCCCT CCT CCGCCCCT CCT CCGCCCCT CCT CCGCCCCC CCT CCGCCCCC CCT CCCCCCCC CCT CCCCCCCC CCT CCCCCCCC	CTA CGCGTTAGAG CTCGCTGATC GAT GCGCAATCTC GAGCGACTAG GCC ATCTGTTGTT TGCCCCTCCC GGG TAGACAACAA ACGGGGAGGG CCA CTCCCACTGT CCTTTCCTAA GGT GAGGGTGACA GGAAAGGATT CTG AGTAGGTGTC ATTCTATTCT GAC TCATCCACAG TAAGATAAGA CCC CCTCCTAACC CTTCTGTTAT CCC TCGCTCACTG ACTCGCTGCG AGG AGCGATGAC TGAGCGACGC ATC AGCTCACTCA AAGGCGGTAA TAG TCGAGTGAGT TTCCGCCATT ACG CAGGAAAGAA CATGTGAGCA ACG CAGGAAAGAA CATGTGAGCA ACG CAGGAAAGAA CATGTGAGCA ACG CAGGAAAGAA CATGTGAGCA CGT TTCCGGCGGT TGCTGGCGTT TTCCGGCGCA ACGACCGCAA CGA TCACAAAAAT CGACGCTCAA CGA TTTCTATGGT CCGCAAAAGGG CTT CCGACCCTGC CGCTTACCGG CAAA GGCTGGGCCT TCTCAATGCT TTC CGACCCTGC CGCTTACCGG CAAG CGTGGCGCTT TCTCAATGCT TTC CGACCCTGC CAAGCTGGGC GAC CGCTGCGCAT TCTCAATGCT TTC GCACCCTGC CAAGCTGGGC GAC CGCTGCGCCT TATCCGGTAA CGC CGCTGCGCCT TATCCGGTAA CGC CGCTGCGCCT TATCCGGTAA CGC CGCTGCGCCT TATCCGGTAA CCG CGCACGCGGA ATAGGCCATT CACA CGACTTATCG CCACTGGCAG CCAC CGCACGCGGA ATAGGCCATT CACA CGACTTATCG CCACTGGCAG CCAC CGCACTTATCG CCACTGGCAG CCACCCCACACTGCAC CCACTGCCACTTATCG CCACTGGCAG CCACCCCCACACACACACACACACACACACACAC	CTA CGCGTTAGAG CTCGCTGATC AGCCTCGACT GCGCAATCTC GAGCGACTAG TCGGAGCTGA GCGCAATCTC GAGCGACTAG TCGGAGCTGAC GCGCTACCC CCGTGCCTTC CCG TAGACAACAA ACGGGGAGGG GGCACGGAAG CCA CTCCCACTGT CCTTTCCTAA TAAAATGAGG GGAAGGATT ATTTACTCC CTG AGTAGGTGACA GGAAAGGATT ATTTACTCC CTG AGTAGGTGACA GAAGACAATA GCAGGCATGC CCCCCCACCC CCC CCCCCCCCC CCC CCCCCCCC	CTA CCCCACCA ACCCCCACC CACCCCCCCCCCCCCC

pCMV-NS34A

4851	AGCAGAGCGA TCGTCTCGCT	GGTATGTAGG CCATACATCC	CGGTGCTACA GCCACGATGT	GAGTTCTTGA CTCAAGAACT	AGTGGTGGCC TCACCACCGG	
4901	TAACTACGGC ATTGATGCCG	TACACTAGAA ATGTGATCTT	GGACAGTATT CCTGTCATAA	TGGTATCTGC ACCATAGACG	GCTCTGCTGA CGAGACGACT	
4951	AGCCAGTTAC TCGGTCAATG	CTTCGGAAAA GAAGCCTTTT	AGAGTTGGTA TCTCAACCAT	GCTCTTGATC CGAGAACTAG	CGGCAAACAA GCCGTTTGTT	
5001	ACCACCGCTG TGGTGGCGAC	GTAGCGGTGG CATCGCCACC	TTTTTTTGTT AAAAAAACAA	TGCAAGCAGC ACGTTCGTCG	AGATTACGCG TCTAATGCGC	
5051	CAGAAAAAA GTCTTTTTT	GGATCTCAAG CCTAGAGTTC	AAGATCCTTT TTCTAGGAAA	GATCTTTTCT CTAGAAAAGA	ACGGGGTCTG TGCCCCAGAC	
5101	ACGCTCAGTG TGCGAGTCAC	GAACGAAAAC CTTGCTTTTG	TCACGTTAAG AGTGCAATTC	GGATTTTGGT CCTAAAACCA	CATGAGATTA GTACTCTAAT	
5151	TCAAAAAGGA AGTTTTTCCT	TCTTCACCTA AGAAGTGGAT	GATCCTTTTA CTAGGAAAAT	AATTAAAAAT TTAATTTTA	GAAGTTTTAA CTTCAAAATT	
5201	ATCAATCTAA TAGTTAGATT	AGTATATATG TCATATATAC	AGTAAACTTG TCATTTGAAC	GTCTGACAGT CAGACTGTCA	TACCAATGCT ATGGTTACGA	
5251	TAATCAGTGA ATTAGTCACT	GGCACCTATC CCGTGGATAG	TCAGCGATCT AGTCGCTAGA	GTCTATTTCG CAGATAAAGC	TTCATCCATA AAGTAGGTAT	
5301	GTTGCCTGAC CAACGGACTG			ACGATACGGG TGCTATGCCC		
5351	ATCTGGCCCC TAGACCGGGG			AGACCCACGC TCTGGGTGCG		
5401	CAGATTTATO GTCTAAATAG			GAAGGGCCGA CTTCCCGGCT		
5451				TCTATTAATT AGATAATTAA	GTTGCCGGGA CAACGGCCCT	
5501	AGCTAGAGTA TCGATCTCAT	AGTAGTTCGC TCATCAAGCG	CAGTTAATAG GTCAATTATC	TTTGCGCAAC AAACGCGTTG	GTTGTTGCCA CAACAACGGT	
5551					GGCTTCATTC CCGAAGTAAG	
5601	AGCTCCGGTT TCGAGGCCAA	CCCAACGATC GGGTTGCTAG	AAGGCGAGTI TTCCGCTCAA	ACATGATCCC TGTACTAGGG	CCATGTTGTG GGTACAACAC	
5651					AGAAGTAAGT TCTTCATTCA	
5701					TAATTCTCTT ATTAAGAGAA	
5751					AGTACTCAAC TCATGAGTTG	
				-		•

pCMV-NS34A

5801	CAAGTCATTC GTTCAGTAAG	TGAGAATAGT ACTCTTATCA	GTATGCGGCG CATACGCCGC	ACCGAGTTGC TGGCTCAACG	TCTTGCCCGG AGAACGGGCC	
5851	CGTCAATACG GCAGTTATGC	GGATAATACC CCTATTATGG	GCGCCACATA CGCGGTGTAT	GCAGAACTTT CGTCTTGAAA	AAAAGTGCTC TTTTCACGAG	
5901	ATCATTGGAA TAGTAACCTT	AACGTTCTTC TTGCAAGAAG	GGGGCGAAAA CCCCGCTTTT	CTCTCAAGGA GAGAGTTCCT	TCTTACCGCT AGAATGGCGA	
5951	GTTGAGATCC CAACTCTAGG	AGTTCGATGT TCAAGCTACA	AACCCACTCG TTGGGTGAGC	TGCACCCAAC ACGTGGGTTG	TGATCTTCAG ACTAGAAGTC	
6001	CATCTTTTAC GTAGAAAATG	TTTCACCAGC AAAGTGGTCG	GTTTCTGGGT CAAAGACCCA	GAGCAAAAAC CTCGTTTTTG	AGGAAGGCAA TCCTTCCGTT	
6051	AATGCCGCAA TTACGGCGTT	AAAAGGGAAT TTTTCCCTTA	AAGGGCGACA TTCCCGCTGT	CGGAAATGTT GCCTTTACAA	GAATACTCAT CTTATGAGTA	
6101					TATTGTCTCA ATAACAGAGT	
6151					AATAGGGGTT TTATCCCCAA	
6201					AAACCATTAT TTTGGTAATA	
6251	TATCATGACA ATAGTACTGT	TTAACCTATA AATTGGATAT	AAAATAGGCG	TATCACGAGG	CCCTTTCGTC GGGAAAGCAG	

FIGURE 10

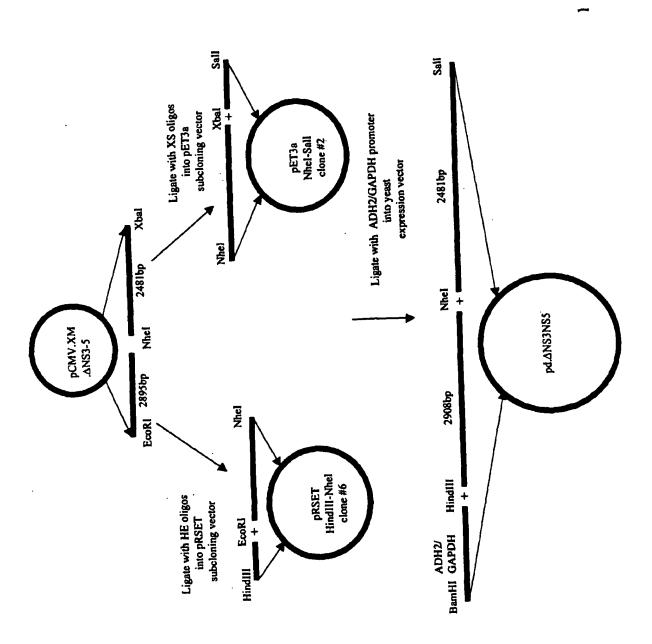


Diagram 1

- LeuAsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGly
 62 CTCAACCCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGG
 GAGTTGGGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCC
- IleAspProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyr
 122 ATCGATCCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTAC
 TAGCTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATG

 122 CLAI,
- SerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIle
 182 TCCACCTACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGGCGCTTATGACATAATA
 AGGTGGATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTAT
- IleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeu
 242 ATTTGTGACGAGTGCCACTCCACGGATGCCACTCCTTTGGCCATTGGCACTGCCTT
 TAAACACTGCTCACGGTGAGGTGCCTACGGTGAGGTAGAACCCGTAACCGTGACAGGAA
- AspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGly
 302 GACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCCCACCGCCACCCCTCCGGGC
 CTGGTTCGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCG
 309 ALWN1,
- SerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIle
 362 TCCGTCACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATC
 AGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAG
- CyshisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsn
 482 TGTCATTCAAAGAAGAAGTGCGACGAACTCGGCGCAAAGCTGGTCGCATTGGGCATCAAT
 ACAGTAAGTTTCTTCTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTA
- AlavalAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValVal
 542 GCCGTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTC
 CGGCACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAG
 556 SAC2, 566 DRD1,
- . ValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAsp 602 GTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGAC CAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTG
 - 621 BSPH1,
 - ${\tt CysAsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGlu}$

- 662 TGCAATACGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAG
 ACGTTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTC
- ThrileThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArg
 722 ACAATCACGCTCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGG
 TGTTAGTGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCC
- GlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAsp
 782 GGGAAGCCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGAC
 CCCTTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGAGGCCGTACAAGCTG
 - 822 BGLI, 839 DRD1,
- - 887 SACI.
- GluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAsp 902 GAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGAC CTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTG
 - 937 SMAI XMAI,
- - 991 STUI,
- SerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrVal
 TCCCAGACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTG
 AGGGTCTGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCAC
 - 1075 DRA3,
- CysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArg
 1082 TGCGCTAGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGC
 ACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCG
- LeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsn
 1142 CTCAAGCCCACCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAAT
 GAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTA
 - 1156 NCOI,
- - 1236 BSPH1, 1240 DRD1, 1243 AVA3, 1251 EAG1 XMA3, 1256 DRD1,
- GluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyr
 1262 GAGGTCGTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTTGGCTGCTGTTTGGCCGCGTAT
 CTCCAGCAGTGCTCGTGGACCCACGAGCCACCGCAGGACCGACGAAACCGGCGCATA

gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1115 1110 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551

FIGURE 11 - Page 3

CysL uSerThrGlyCysValValIl ValGlyArgValValLeuSerGlyLysProAla 1322 TGCCTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCA ACGGACAGTTGTCCGACCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGT

1375 NAEI,

IleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGln
1382 ATCATACCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAG
TAGTATGGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTC

1391 DRD1,

- HisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeu
 1442 CACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTC
 GTGAATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAG
- GlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsn
 GGCCTCCTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAAC
 CCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTG

1508 PSTI, 1513 TTH3I,

TrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGln
1562 TGGCAAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAA
ACCGTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTT

1571 XHOI, 1592 NDEI,

TyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPhe
1622 TACTTGGCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTT
ATGAACCGCCCGAACAGTTGCGACGACCATTGGGGCGGTAACGAAGTAACTACCGAAAA

1649 BSTE2,

ThralaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGly
1682 ACAGCTGCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGG
TGTCGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCC

1683 ALWN1 PVU2,

GlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGly
1742 GGGTGGGTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGC
CCCACCCACCGACGGCTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCG

1800 ESP1,

LeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAla
1802 TTAGCTGGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCA
AATCGACCGCGGGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGT

1808 KAS1 NARI,

GlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluVal
1862 GGGTATGGCGCGGGGGGGGGGGGGGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTC
CCCATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAG

FIGURE 11 - Page 4

1884 SACI, 1905 BSPH1,

ProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuVal
1922 CCCTCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCTCGTA
GGGAGGTGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCAT

1934 TTH3I,

ValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaVal
1982 GTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGGAGTG
CAGCCGCACCAGACACGTCGTTATGACGCGGCCGTCCACCGTCAC

2010 NAEI, 2023 SMAI XMAI,

GlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHis
CAGTGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCAC
GTCACCTACTTGGCCGACTATCGGAAGCGGAGGCCCCCCTTGGTACAAAGGGGGTGCGTG

2073 SMAI XMAI, 2099 DRA3,

2165 ALWN1, 2170 MST2.

TyrValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrVal
TACGTGCCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTA
ATGCACGGCCTCTCGCTACGTCGACGGCGCGCAGTGACGTTAGAGTCGTCGGAGTGACAT

2121 PVU2,

- ThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSer
 2162 ACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCC
 TGGGTCGAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGG
 - GlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThr

2222 GGTTCCTGGCTAAGGGACATCTGGGACTGATATGCGCTCCACAACTCGCTGAAATTCTGG
CCAAGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGG

2226 ECON1,

2291 ESP1, 2306 PVU2, 2316 BAMHI,

- GlyTyrLysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAla
 2342 GGGTATAAGGGGGTCTGGCGAGGGGACGCATCATGCACACTCGCCACTGTGGAGCT
 CCCATATTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGA
- GluIleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArg
 2402 GAGATCACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGG
 CTCTAGTGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCC

2431 BSAB1, 2447 AVR2, 2454 SSE83871, 2455 PSTI,

AsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeu
AACATGTGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTT
TTGTACACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAA

FIGURE 11 - Page 5

2486 ASE1,	2503	APAI,	•
------------	------	-------	---

- ProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIle
 CCTGCGCCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATA
 GGACGCGGCTTGATGCAAGCGCGATACCTCCACAGACGTCTCCTTATGCACCTCTAT
 - 2559 PSTI,

2600 DRA3,

- ArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysPro
 2582 AGGCAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCCG
 TCCGTCCACCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGC
- CysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPhe
 TGCCAGGTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTT
 ACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAA
- AlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGlu
 CGCGCCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAA
 CGCGGGGGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTT
- TyrProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSer

 TACCCGGTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCC
 ATGGGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGG

 2763 HGIE2, 2815 AAT2,
- MetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGly
 2822 ATGCTCACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGA
 TACGAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCT
 - 2856 EAG1 XMA3,
- SerProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAla
 2882 TCACCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCA
 AGTGGGGGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGT
 - 2895 BALI, 2909 NHEI,
- ThrCysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrp
 ACTTGCACCGCTAACCATGACTCCCCTGATGCTGAGGCTCATAGAGGCCAACCTCCTATGG
 TGAACGTGGCGATTGGTACTGAGGGGACTACCGAGTATCTCCGGTTGGAGGATACC
 - 2972 ESP1, 2975 SACI,
- ArgGlnGIuMētGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeu
 3002 AGGCAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTG
 TCCGTCCTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGAC
- AspSerPheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGlu
 3062 GACTCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGGGGGGAGATCTCCGTACCCGCAGAA
 CTGAGGAAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTT
 - 3102 BGL2,

- - 3149 ALWN1, 3170 EAG1 XMA3,
- AsnProProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGly
 3182 AACCCCCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGC
 TTGGGGGGGGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCG
 - 3223 HGIE2, 3235 NCOI,
- CysProLeuProProProLysSerProProValProProProArgLysLysArgThrVal
 3242 TGCCCGCTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCAAGAAGCCGAAGAGCGGACGCTGCAC
 ACGGGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCAC
- ValleuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGly
 3302 GTCCTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGC
 CAGGAGTGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCG
 - 3338 SACI, 3352 HIND3,
- SerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGly
 3422 TCTGGCTGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAGGGG
 AGACCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCC
 - 3443 EAM11051,
- GluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsn
 3482 GAGCCTGGGGATCCTGGATCTTAGCGACGGTCATGGTCAACGGTCAGTAGTGAGGCCAAC
 CTCGGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTG
 - 3490 BAMHI, 3491 BSAB1, 3493 BSPE1,
- AlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrPro 3542 GCGGAGGATGTCGTGCTCCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCG CGCCTCCTACAGCACGACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGC
 - 3595 DRA3,
- CysAlaAlaGluGluGlnLysLeuProlleAsnAlaLeuSerAsnSerLeuLeuArgHis 3602 TGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCAC ACGCGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTG
 - 3606 SAC2, 3617 ALWN1, 3661 PFLM1,
- HisAsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThr
 CACAATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACA
 GTGTTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGT
 - 3687 DRA3,
 - PheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValL uLysGluValLysAla

FIGURE 11 - Page 7

- 3722 TTTGACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCA
 AAACTGTCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGT
- AlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrPro
 GCGGCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCC
 CGCCGCAGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGG

3822 HIND3,

ProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArg
3842 CCACACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGA
GGTGTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCT

3881 AAT2, 3896 BGLI,

- LysalavalThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrPro
 3902 AAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCA
 TTCCGGCATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGT
- IleAspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGly
 3962 ATAGACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGGT
 TATCTGTGATGGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCA
- ArgLysProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMet
 4022 CGTAAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATG
 GCATTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTAC
- AlaLeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPhe
 4082 GCTTTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTC
 CGAAACATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAG
- GIntyrserProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThr
 4142 CAATACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACC
 GTTATGAGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGG

4166 ECORI,

- ProMetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIle
 CCAATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATC
 GGTTACCCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAG
 - 4235 DRD1, 4242 ALWN1,
- ArgThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIle
 4262 CGTACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATC
 GCATGCCTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAG
 - 4307 BGLI, 4314 BALI,
- LysSerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsn
 4322 AAGTCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAAC
 TTCAGGGAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTG

4351 APAI,

CysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeu
4382 TGCGGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTC

haGURE 11 - Page 8

ACGCCGATAGCGTCCACGGCGCGCCCCCCCATGACTGTTGATCGACACCATTGTGGGAG

ThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMet

4442 ACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATG
TGAACGATGTAGTTCCGGGCCCGTCGGACAGCTCCGAGGTCCTGACGTGGTAC

4458 SMAI XMAI.

LeuValCysGlyAspAspLeuValVallleCysGluSerAlaGlyValGlnGluAspAla
4502 CTCGTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCG
GAGCACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGC

4514 DRD1, 4517 TTH3I,

- ProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAla
 4622 CCACAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCC
 GGTGTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGG

4643 SACI,

HisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAla
4682 CACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCTCGCG
GTGCTGCCGCGACCTTTCTCCCAGATGATGGAGTGGCACTGGGATGTTGGGGGGAGCGC

4737 NRUI,

- ArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIle
 4742 AGAGCTGCGTGGGAGACAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATC
 TCTCGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAG
- MetPheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeu
 4802 ATGTTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTT
 TACAAACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAA

4812 PFLM1, 4813 DRA3,

IleAlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSer
4862 ATAGCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCC
TATCGGTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGG

4899 BGL2, .

IleGluProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSer
4922 ATAGAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCA
TATCTTGGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGT

4960 NCOI,

LeuHisSerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGly
4982 CTCCACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGG
GAGGTGTCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCC

5021 SPHI, 5041 KPNI,

- ValProProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgL uLeuAla 5042 GTACCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCC CATGGCGGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGG
 - 5070 APAI, 5097 BALI,
- ArgGlyGlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLys
 5102 AGAGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAG
 TCTCCTCCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTC
 - 5119 NDEI,
- LeuLysLeuThrProIleAlaAlaAlaGlyGlnLeuAspLeuSerGlyTrpPheThrAla
 5162 CTCAAACTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCT
 GAGTTTGAGTGAGGTTATCGCCGGCGACCGTCGACCTGAACAGGCCGACCAAGTGCCGA
 - 5180 NOTI, 5181 EAG1 XMA3, 5188 BALI, 5192 PVU2,
- GlyTyrSerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrp
 5222 GGCTACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGCTGGATCTGG
 CCGATGTCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGGCGACCTAGACC
 - 5246 DRA3,
- PheCysLeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgOP

 5282 TTTTGCCTACTCCTGCTGCAGGGGTAGGCATCTACCTCCCCAACCGATGAAGG
 AAAACGGATGAGGACGACGTCCCCATCCGTAGATGGAGGGGGTTGGCTACTTCC
 - 5301 PSTI, 5331 HGIE2,
- - '5378 XBAI, 5390 SALI,

FIGURE 12

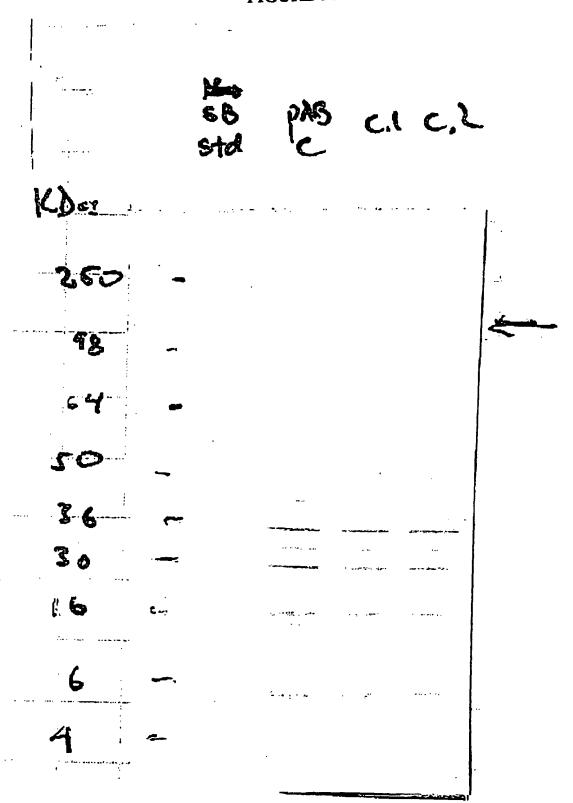


FIGURE 13

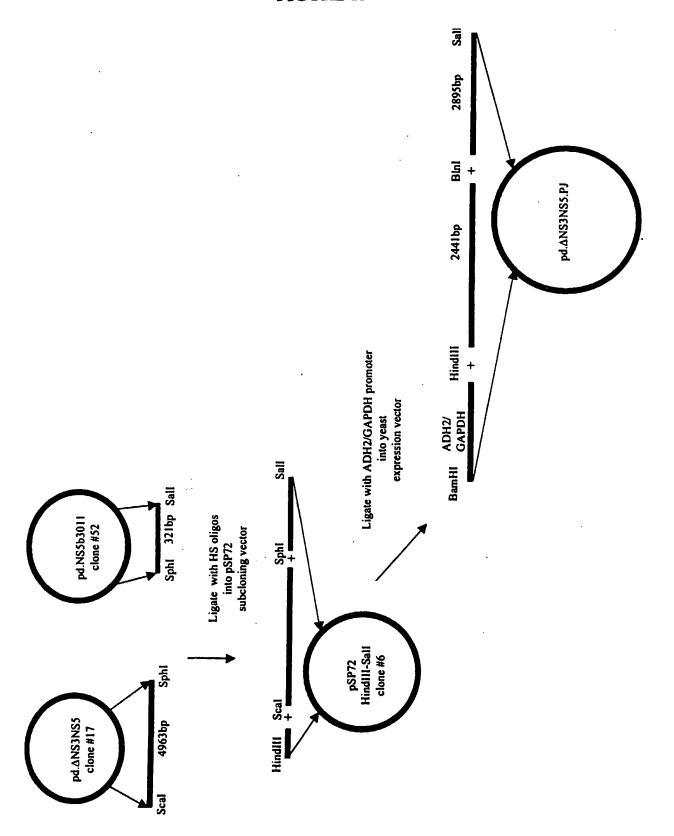


FIGURE 14 - Page 1

- - 1 HIND3, 24 NDEI, 52 SCAI,
- ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp 62 CCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 - 116 CLAI,
- ProAsmileArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG
- TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCCTTGCCGACGGCGGTGCTCGGGGGGGCGCTTATGACATAATAATTTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACA
- AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 GACGAGTGCCACTCCACGGATGCCACTCTTGGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACTGGTT
- AlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerVal
 302 GCAGAGACTGCGGGGGGGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAG
 303 ALWN1.
- ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTGCTCTGTCCACCACCGGAGAGATCCCTTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA
- TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTA
- SerLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 482 TCAAAGAAGAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 AGTTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCAC
- AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal
 542 GCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCAC
 - 550 SAC2, 560 DRD1,
- AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVallleAspCysAsn
 GCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAAT
 CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTA
 - 615 BSPH1.

FIGURE 14 - Page 2

- 662 ACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG
- ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 722 ACGCTCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAG
 TGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCTTC
- ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTTGTGGCACCGGGGGGGGGCCCCTCCGGCATGTTCGACTCGTCC
 GGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGAGGCCGTACAAGCTGAGCAGC
 - 816 BGLI, 833 DRD1,
- - 881 SACI,
- ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAA
 - 931 SMAI XMAI,
- GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln GAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAG CTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC
 - 985 STUI,
- ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
 1022 ACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
 TGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGA
 - 1069 DRA3,
- ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
 1082 AGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAG
 TCCCGAGTTCGGGGAGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC
- ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG
 - 1150 NCOI,
 - - 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
 - ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGAAACCGGCGCATAACGGAC

FIGURE 14 - Page 3

SerThrGlyCysValValIleVålGlyArgValValLeuSerGlyLysProAlaTieTle
1322 TCAACAGGCTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATA
AGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTAT

1369 NAEI,

ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
1382 CCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACCTTCTCACGAGAGTCGTGAAT

1385 DRD1,

- ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTC
 GGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAG
- LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTT

 1502 PSTI, 1507 TTH3I,
- LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTG
 TTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAAC

1565 XHOI, 1586 NDEI,

- AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 GCGGGCTTGTCAACGCTGCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCT
 CGCCCGAACAGTTGCGACGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGA
 - 1643 BSTE2, 1677 ALWN1 PVU2,
- AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp

 GCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGG

 CGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACC
- ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 1742 GTGGCTGCCCAGCTCGCCCCCGGTGCCGTACTGCCTTTGTGGGCGCTGGCTTAGCT
 CACCGACGGTCGAGCGGGGGGCCACGGCGATGACGAAACACCCCGCGACCGAATCGA

1794 ESP1,

1802 KAS1 NARI,

- GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
 1802 GGCGCCGCCATCGGCAGTGTTGGACCTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTAT
 CCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATA
- GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
 1862 GGCGCGGGCGGGGGGGGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCC
 CCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGG

1878 SACI, 1899 BSPH1.

FIGURE 14 - Page 4

- ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly

 1922 ACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGC
 TGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCG
 - ≥ 1928 TTH3I,
- ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
 1982 GTGGTCTGTGCAGCAATACTGCGCCGGCCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGG
 CACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACC
 - 2004 NAEI, 2017 SMAI XMAI,
- MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
 ATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTG
 TACTTGGCCGACTATCGGAAGCGGAGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCAC
 - 2067 SMAI XMAI, 2093 DRA3,
- ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
 CCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAG
 GGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC
 - 2115 PVU2, 2159 ALWN1,
- LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrTnrProCysSerGlySer
 2162 CTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
 GAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAAG
 - 2164 MST2, 2220 ECON1,
- TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
 ACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGAT
- - 2285 ESP1, 2300 PVU2, 2310 BAMHI,
- LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 2342 AAGGGGGTCTGGCGAGGGGACGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG
- ThrGlyHisValLysAsnGlyThrMetArglleValGlyProArgThrCysArgAsnMet
 2402 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC
 - 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,
- TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 TGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGC
 - 2480 ASE1, 2497 APAI,
 - ${\tt ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln}$

FIGURE 14 - Page 5

2522 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAGGGCTTGATGTGCAAGCGCGGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC

2553 PSTI,

ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
2582 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAG
CACCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGCACGGTC

2594 DRA3,

- ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro 2642 GTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCC CAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGG
- ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCG
 GGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2,

ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
2762 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTC
CATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAG

2809 AAT2,

ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
2922 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCC
TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGG

2850 EAG1 XMA3,

ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGC
GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG

2889 BALI, 2903 NHEI,

ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln
2942 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
TGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTC

2966 ESP1, 2969 SACI,

- GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGG
- PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 3062 TTCGATCCGCTTGTGGCGGAGGAGGAGGAGGGGGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGAC

3096 BGL2.

ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro 3122 CGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCC

FIGURE 14 - Page 6

GCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCGCCGGCCTGATATTGGGG

- 3143 ALWN1, 3164 EAG1 XMA3,
- ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 - 3217 HGIE2, 3229 NCOI,
- LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAG
- ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCC
 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 - 3332 SACI, 3346 HIND3,
- CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGA
 - 3437 EAM11051.
- GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTC
 - 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
- AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGG
 - 3589 DRA3, 3600 SAC2,
- AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602 GCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAAT
 CGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTA
 - 3611 ALWN1, 3655 PFLM1,
- LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 3662 TTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGAC
 AACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTG
 - 3681 DRA3,
- ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCG
 TCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGC

FIGURE 14 - Page 7

SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis 3782 TCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCACAC AGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTG 3816 HIND3, ${\tt SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla}$ 3942 TCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC AGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGG 3875 AAT2, 3890 BGLI, ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp 3902 GTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGAC CATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG ${\tt ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys}$ 3962 ACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGGTCGTAAG TGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTC ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu 4022 CCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTG GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAAC TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr 4082 TACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC ATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATG ${\tt SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet}$ 4142 TCACCAGGACAGCGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATG AGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTAC 4160 ECORI, ${\tt GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr}$ 4202 GGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACG CCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC 4229 DRD1, 4236 ALWN1,

- GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer

 4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
 CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGG
 - 4301 BGLI, 4308 BALI,
- LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 4322 CTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGC
 GAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCG
 - 4345 APAI,
- TyrargargCysargalaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCCCCCCCATGACTGTTGATCGACACCATTGTGGGAGTGAACG

FIGURE 14 - Page 8

TyrileLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
4442 TACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
ATGTAGTTCCGGGCCCGTCGGACAGCTCCGGCGTCCCGAGGTCCTGACGTGGTACGAGCAC

. 4452 SMAI XMAI,

CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGGGGAGC
ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCG

4508 DRD1, 4511 TTH3I,

- LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGGTGTT
- ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG

4637 SACI,

GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
4682 GGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGAGCT
CCGCGACCTTTCTCCCAGATGATGGAGTGGCACTGGGATGTTGGGGGGAGCGCTCTCGA

4731 NRUI,

- AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe 4742 GCGTGGGAGACAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTT CGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAA
- AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
TCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTT

4893 BGL2,

ProLeuAspLeuProProllelleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCAC GGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG

4954 NCOI,

SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCG
TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGC

5015 SPHI, 5035 KPNI,

ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly

FIGURE 14 - Page 9

- 5042 CCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGAGGA GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT
 - 5064 APAI, 5091 BALI,
- GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
 5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAA
 CCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTT
 - 5113 NDEI,
- - 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
- SerGlyGlyAsplleTyrHisSerValSerHisAlaArgProArgTrplleTrpPheCys
 5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGGTGGATCTGGTTTTGC
 TCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGGCGACCTAGACCAAAACG
 - 5240 DRA3,
- LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgOP

 5282 CTACTCCTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCGATGAATAGTCGAC
 GATGAGGACGACGACGTCCCCATCCGTAGATGGAGGGGGTTGGCTACTTATCAGCTG

.. 1 -

5295 PSTI, 5336 SALI,

FIGURE 15



FIGURE 16 - Page 1

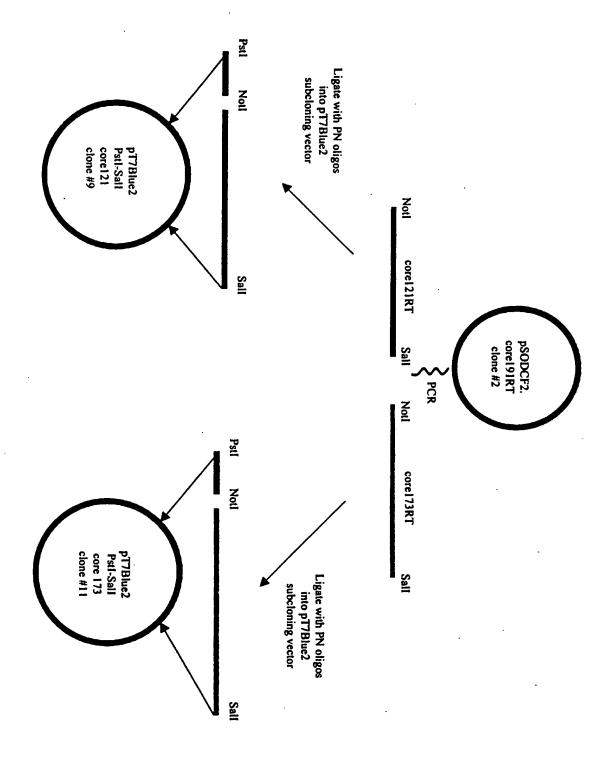


FIGURE 16 - Pa 2

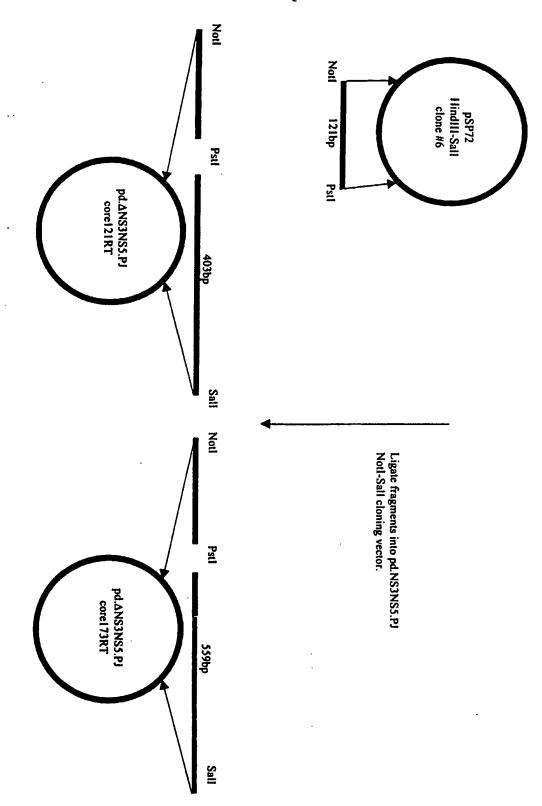


FIGURE 17 - Page 1

- - 1 HIND3, 24 NDEI, 52 SCAI,
- ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp
 62 CCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 - 116 CLAI,
- ProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG
- TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCCTTGCCGACGGCGGTGCTCGGGGGGCGCTTATGACATAATATTTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACA
- AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACTGGTT
- AlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerVal
 GCAGAGACTGCGGGGGGGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAG
 303 ALWN1,
- ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA
- TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTA

FIGURE 17 - Page 2

- SerLysLysLysCysAspGluL uAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 TCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 AGTTTCTTCTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCAC
- AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal
 542 GCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCAC

550 SAC2, 560 DRD1,

AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsn 602 GCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAAT CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGACCCACTATCTGACGTTA

615 BSPH1,

- ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 662 ACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTTCACCATTGAGACAATC
 TGCACACAGTGGGTCTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG
- ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAG
 TGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCTTC
- ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCC
 GGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGAGGCCGTACAAGCTGAGCAGG

816 BGLI, 833 DRD1,

881 SACI.

- ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAA
 - 931 SMAI XMAI,
- GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln 962 GAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAG CTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC

985 STUI,

ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
1022 ACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
TGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGA

1069 DRA3,

ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
1082 AGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAG

FIGURE 17 - Page 3

TCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC

- ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG
 - 1150 NCOI,
- - 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
- ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAACCGGCGCATAACGGAC
- SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGGAAGCCGGCAATCATA
 AGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTAT
 - 1369 NAEI,
- ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACCTTCTCACGAGAGTCGTGAAT

 1385 DRD1,
- ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTC
 GGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAG
- LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTT

 1502 PSTI, 1507 TTH3I,
- LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
 1562 AAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTG
 TTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAAC
 1565 XHOI, 1586 NDEI,
 - AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
 1622 GCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCT
 CGCCCGAACAGTTGCGACGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGA
 - 1643 BSTE2, 1677 ALWN1 PVU2,
 - AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGG
 CGACAGTGGTCGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACC

FIGURE 17 - Page +

ValalaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
1742 GTGGCTGCCCAGCTCGCCCCCGGTGCCGTACTGCCTTTGTGGGCGCTGGCTTAGCT
CACCGACGGTCGAGCGGGGGGCCACGGATGACGGAAACACCCGGGACCGAATCGA

. 1794 ESP1,

GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr

GGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTAT

CCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATA

1802 KAS1 NARI,

1878 SACI, 1899 BSPH1,

ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
1922 ACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGC
TGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCG

1928 TTH3I,

ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
1982 GTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGG
CACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACC

2004 NAEI, 2017 SMAI XMAI,

MetAsnArgLeulleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
ATGAACCGGCTGATAGCCTTCGCCTCCGGGGGAACCATGTTTCCCCCACGCACTACGTG
TACTTGGCCGACTATCGGAAGCGGAGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCAC

2067 SMAI XMAI, 2093 DRA3,

ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
CCGGAGAGCGATGCAGCTGCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAG
GGCCTCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC

2115 PVU2, 2159 ALWN1,

LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
2162 CTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
GAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAGG

2164 MST2, 2220 ECON1,

- TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 2222 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
 ACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGAT

2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 17 - Page 5

- LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 2342 AAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG
- ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC
 - 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,
- TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla 2462 TGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCG ACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGC
 - 2480 ASE1, 2497 APAI,
- ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC
 - 2553 PSTI,
- ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
 2582 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAG
 CACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGCACGGTC
 - 2594 DRA3,
- ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro 2642 GTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCC CAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGG
- ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCG
 GGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC
 - 2757 HGIE2,
- ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 2762 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTC
 CATCCCAGCGTTAATGGAACGCTCGGGCTTGCCCTGCACCGGCACAACTGCAGGTACGAG
 - 2809 AAT2,
- ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 2822 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGG
 - 2850 EAG1 XMA3,
- ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGC
 GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG
 - 2889 BALI, 2903 NHEI,

FIGURE 17 - Page 6

- ThrAlaAsnHisAspSerProAspAlaGluLeuIl GluAlaAsnLeuLeuTrpArgGln
 2942 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCCGAGTATCTCCGGTTGGAGGATACCTCCGTC
 - _ 2966 ESP1, 2969 SACI,
- GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 GAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGG
- PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGGGGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGAC
 - 3096 BGL2,
- - 3143 ALWN1, 3164 EAG1 XMA3,
- ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 - 3217 HGIE2, 3229 NCOI,
- LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCAAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAG
- ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCC
 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 - 3332 SACI, 3346 HIND3,
- CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGA
 - 3437 EAM11051,
- GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTC
 - 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
- AspValValCysCysSerMetSerTyrS rTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGG

FIGURE 17 - Page 7

3589 DRA3, 3600 SAC2, -

- AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602 GCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAAT
 CGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTA
 - 3611 ALWN1, 3655 PFLM1,
- LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 3662 TTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGAC
 AACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTG
- ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCG
 TCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGC
- SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 TCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCACAC
 AGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTG
 - 3816 HIND3,

3681 DRA3,

- SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 TCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC
 AGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGG
 - 3875 AAT2, 3890 BGLI,
- ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG
- ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 3962 ACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTC
- ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTG
 GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAAC
- TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC
 ATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATG
- SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 TCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATG
 AGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTAC
 - 4160 ECORI,
- GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC

FIGURE 17 - Page 8

4229 DRD1, 4236 ALWN1,

GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGG

4301 BGLI, 4308 BALI,

LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
4322 CTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGC
GAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCG

4345 APAI,

- TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCGCGGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACG
- TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 4442 TACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCCGTCGGACAGCTCCCGAGGTCCTGACGTGGTACGAGCAC

4452 SMAI XMAI,

- CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGGTCCAGGAGGACGCGGCGAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCG
 - 4508 DRD1, 4511 TTH3I,
- LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGGTGTT
- ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG

4637 SACI,

GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
4682 GGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGAGCT
CCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCTCGA

4731 NRUI,

- AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe 4742 GCGTGGGAGACAGCAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTT CGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAA
- AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu I le Tyr Gly Ala Cys Tyr Ser I le Glu

FIGURE 17 - Page 9

- 4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
 TCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTT
 - 4893 BGL2,
- ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCAC
 GGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG
 - 4954 NCOI,
- SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCG
 TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGC
 - 5015 SPHI, 5035 KPNI,
- ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
 5042 CCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGAGGA
 GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT
 - 5064 APAI, 5091 BALI,
- GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys 5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAA CCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTT
 - . 5113 NDEI,
- - 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
- SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
 5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGCTGGATCTGGTTTTGC
 TCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGGCGACCTAGACCAAAACG
 - 5240 DRA3,
- LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
 CTACTCCTGCTTGCTGCAGGGGTAGGCATCTCCTCCCCCAACCGAATGAGCACGAAT
 GATGAGGACGACGTCCCCATCCGTAGATGAGGAGGGGGTTGGCTTACTCGTGCTTA
 - 5295 PSTI.
 - ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
 5342 CCTAAACCTCAAAGAAAGACCAACCGTAACACCAACCGGCGGCCGCAGGACGTCAAGTTC
 GGATTTGGAGTTTCTTGTTTGCATTGTGGTTGGCCGCCGGCGTCCTGCAGTTCAAG
 - 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMAI XMAI,
 - ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
 5402 CCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTG
 GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAAC

FIGURE 17 - Page 10

5449 APAI,

GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
5462 GGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGTCAGCCT
CCACACGCGCGCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGA

5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,

IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
5522 ATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCC
TAGGGGTTCCGAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGG

5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,

5702 AC TG

FIGURE 18 - Page 1

- - 1 HIND3, 24 NDEI, 52 SCAI,
- ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp 62 CCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 - 116 CLAI,
- ProAsnileArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG
- TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCCTTGCCGACGGCGGTGCTCGGGGGGGCGCTTATGACATAATATTTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACA
- AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCACTCCACGGATGCCACATCCATCTTGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACTGGTT
- AlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerVal
 GCAGAGACTGCGGGGGGGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAG

 303 ALWN1,
- ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGGGAGAGATCCCTTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA
- TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTTCTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTA
- SerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 TCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 AGTTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCAC
- AlaTyrTyrArgGlyLeuAspValSerVallleProThrSerGlyAspValValValVal
 542 GCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCAC
 - 550 SAC2, 560 DRD1,
- AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVallleAspCysAsn
 602 GCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAAT
 CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGACCACTATCTGACGTTA

FIGURE 18 - Page 2

- ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 662 ACGTGTGTCACCCAGACAGTCGATTCAGCCTTGACCCTTCACCATTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG
- ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAG
 TGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCTTC
- ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTTGTGGCACCGGGGGGGGGCCCCTCCGGCATGTTCGACTCGTCC
 GGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGG

816 BGLI, 833 DRD1,

881 SACI.

ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
902 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTT
TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAA

931 SMAI XMAI,

GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln 962 GAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAG CTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC

985 STUI,

ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
1022 ACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
TGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGA

1069 DRA3,

- ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeulleArgLeuLys
 1082 AGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAG
 TCCCGAGTTCGGGGAGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC
- ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG

 1150 NCOI.
- - 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
- ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTG

FIGURE 18 - Page 3

CAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACGGAC

SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaileile
1322 TCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATA
AGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTAT

1369 NAEI,

ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
1382 CCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACCTTCTCACGAGAGTCGTGAAT

1385 DRD1,

- ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTC
 GGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAG
- LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTT
 ^ ^

1502 PSTI, 1507 TTH3I,

LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
1562 AAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTG
TTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAAC

1565 XHOI, 1586 NDEI,

AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla
1622 GCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCT
CGCCCGAACAGTTGCGACGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGA

1643 BSTE2, 1677 ALWN1 PVU2,

- AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGG
 CGACAGTGGTCGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACC
- ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla
 1742 GTGGCTGCCCAGCTCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCT
 CACCGACGGTCGAGCGGGGGGCCACGGCGATGACGCAAACACCCGCGACCGAATCGA

1794 ESP1,

GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr
1802 GGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTAT
CCGCGGCGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATA

1802 KAS1 NARI,

GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
1862 GGCGCGGGCGTGGCGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCC
CCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGG

1878 SACI, 1899 BSPH1,

FIGURE 18 - Page 4

ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
1922 ACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGC
TGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCG

1928 TTH3I,

ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
1982 GTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGG
CACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACC

2004 NAEI, 2017 SMAI XMAI,

MetAsnArgLeulleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
ATGAACCGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTG
TACTTGGCCGACTATCGGAAGCGGAGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCAC

2067 SMAI XMAI, 2093 DRA3,

ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
CCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAG
GGCCTCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC

2115 PVU2, 2159 ALWN1,

LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
2162 CTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
GAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAGG
^

2164 MST2, 2220 ECON1,

- TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
 ACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGAT

2285 ESP1, 2300 PVU2, 2310 BAMHI,

- LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 . 2342 AAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG
 - ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 2402 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC

2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,

TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla

2462 TGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCG
ACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGC

2480 ASE1, 2497 APAI,

FIGURE 18 - Page 5

ProAsnTyrThrPh AlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGin
CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
GGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC

2553 PSTI,

ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
2582 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAG
CACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGCACGGTC

2594 DRA3,

- ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro GTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCC CAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGG
- ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCG
 GGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2,

ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
2762 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTC
CATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAG

2809 AAT2,

ThrAspProSerHislleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
2822 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCC
TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGG

2850 EAG1 XMA3,

ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGC
GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG

2889 BALI, 2903 NHEI,

ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGln
2942 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
TGGCGATTGGTACTGAGGGGACTACGACTCCGATTCTCCGGTTGGAGGATACCTCCGTC

2966 ESP1, 2969 SACI,

- GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGG
- PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGAC

3096 BGL2.

ArgLysSerArgArgPh AlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro

FIGURE 18 - Page 6

- 3122 CGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCC
 GCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGCCAAACCCGCGCCGGCCTGATATTGGGG
 - 3143 ALWN1, 3164 EAG1 XMA3,
- ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 - 3217 HGIE2, 3229 NCOI,
- LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 CTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAG
- ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer

 ACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCC

 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 - 3332 SACI, 3346 HIND3,
- CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGluGluPro
 3422 TGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCTGGAGGGGGAGCCT
 ACGGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGA
 - 3437 EAM11051,
- GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTC
 - 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
- AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla 3542 GATGTCGTGTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCC CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGG
 - 3589 DRA3, 3600 SAC2,
- AlaGluGluGlnLysLeuProlleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 GCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAAT
 CGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTA
 - 3611 ALWN1, 3655 PFLM1,
- LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp TTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGAC AACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTG
 - 3681 DRA3,
- ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla 3722 AGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCG

FIGURE 18 - Page 7

TCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGC

SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
TCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCACAC
AGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTG

3816 HIND3.

SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
TCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC
AGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGG

3875 AAT2, 3890 BGLI,

- ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG
- ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 3962 ACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTC
- ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTG
 GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAAC
- TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC
 ATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATG
- SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 TCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATG
 AGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTAC

4160 ECORI,

- GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC
 - 4229 DRD1, 4236 ALWN1,
- GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
 4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCTGGCCATCAAGTCC
 CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGG
 - 4301 BGLI, 4308 BALI,
- LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
 4322 CTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGC
 GAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCG

4345 APAI,

TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
4382 TATCGCAGGTGCCGCGGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
ATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACG

FIGURE 18 - Page 8

- TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 4442 TACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCCGTCGGACACTCCCGAGGTCCTGACGTGGTACGAGCAC
 - 4452 SMAI XMAI,
- CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
 4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCGAGC
 ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCG
 - 4508 DRD1, 4511 TTH3I,
- LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGGACCCCTGGGGGGTGTT
- ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG
 - 4637 SACI,
- GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 4682 GGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCTCGCGAGAGCT
 CCGCGACCTTTCTCCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCTCGA
 - 4731 NRUI,
- AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 4742 GCGTGGGAGACAGCACACCTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTT
 CGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAA
- AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG
 - 4806 PFLM1, 4807 DRA3,
- ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
 4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
 TCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTT
 - 4893 BGL2,
- ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCAC
 GGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG
 - 4954 NCOI,
- SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCG
 TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGC
 - 5015 SPHI, 5035 KPNI,

FIGURE 18 - Page 9

ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuÂlaArgGiy
5042 CCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTTCTGGCCAGAGGA
GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT

5064 APAI, 5091 BALI,

GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAA
CCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTT

5113 NDEI,

- - 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
- SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
 5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGCTGGATCTGGTTTTGC
 TCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGGCGACCTAGACCAAAACG

5240 DRA3,

LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
5282 CTACTCCTGCTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCGAATGAGCACGAAT
GATGAGGACGAACGACGTCCCCATCCGTAGATGAGGAGGGGGTTGGCTTACTCGTGCTTA

5295 PSTI,

- ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
 5342 CCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGGCGGCCGCAGGACGTCAAGTTC
 GGATTTGGAGTTTCTTCTGGTTTGCATTGTGGTTGGCCGCCGGCGTCCTGCAGTTCAAG
 - 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMAI XMAI,
- ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
 5402 CCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTG
 GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAAC

5449 APAI,

- GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
 5462 GGTGTGCGCGGACGAGAAAGACTTCCGAGCGTCGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGA
 - 5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,
- IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 5522 ATCCCCAAGGCTCGTCGGCCCGAGGGCAGGCCTGGGCTCAGCCCGGGTACCCTTGGCCC
 TAGGGGTTCCGAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGG
 - 5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,
- LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 5582 CTCTATGGCAATGAGGGCTGCGGGTGGCCGGATGGCTCCTGTCTCCCCGTGGCTCTCGG
 GAGATACCGTTACTCCCGACGCCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCC

FIGURE 18 - Page 10

ProSerTrpGlyProThrAspProArgArgArgSerArgAsnLeuGlyLysValIleAsp
CCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGAT
GGATCGACCCCGGGGTGTCTGGGGGCCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTA

5650 APAI, 5696 CLAI,

ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeu

5702 ACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCCCTCTT

TGGGAATGCACGCCGAAGCGGCTGGAGTACCCCCATGTATGGCGAGCAGCCGCGGGGAGAA

5724 HGIE2, 5750 KAS1 NARI, 5756 ECON1,

GlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeuGluAspGlyValAsnTyr 5762 GGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTTCTGGAAGACGGCGTGAACTAT CCTCCGCGACGGTCCCGGGACCGCTACCGCAGGCCCAAGACCTTCTGCCGCACTTGATA

5772 BSTXI, 5775 APAI,

AlaThrGlyAsnLeuProGlyCysSerOC AM
5822 GCAACAGGGAACCTTCCTGGTTGCTCTTAATAGTCGAC
CGTTGTCCCTTGGAAGGACCAACGAGAATTATCAGCTG

5854 SALI,

FIGURE 19

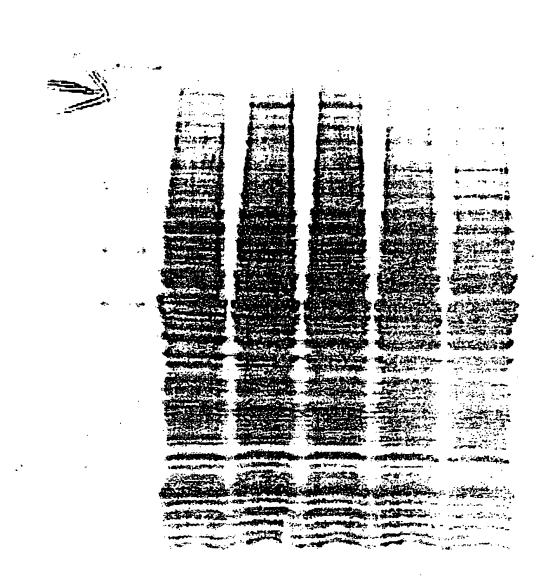


FIGURE 20 - Page 1

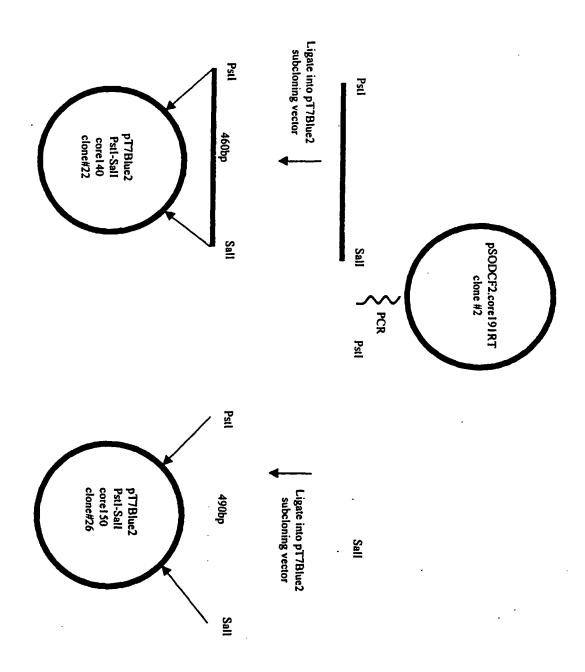


FIGURE 20 - Pas - 2

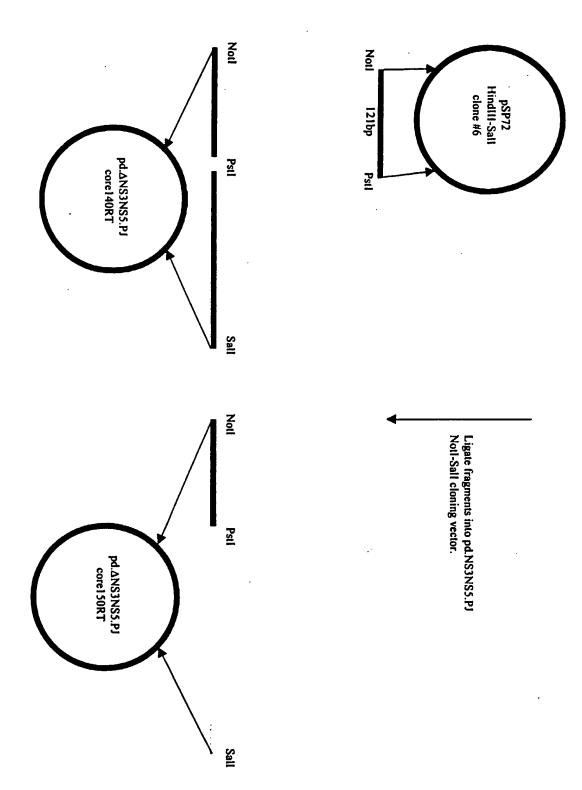


FIGURE 21 - Page 1

- ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp
 62 CCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT
 GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 116 CLAI.
- ProAsnileArgThrGlyValArgThrileThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG
- TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGGCGCTTATGACATAATAATTTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACA
- AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACTGGTT
- ThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA
- TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTTCTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTA

FIGURE 21 - Page 2

- SerLysLysCysAspGluL@uAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 482 TCAAAGAAGAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 .AGTTTCTTCTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCAC
- AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal

 GCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTG

 CGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCAC

550 SAC2, 560 DRD1,

AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVallleAspCysAsn
602 GCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAAT
CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGACGCACTATCTGACGTTA

615 BSPH1,

- ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIle
 662 ACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG
- ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAG
 TGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCCTTC
- ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCTCCGGCATGTTCGACTCGTCC
 GGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGAGGCCGTACAAGCTGAGCAGG

816 BGLI, 833 DRD1,

881 SACI,

- ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAA
 - 931 SMAI XMAI,
- GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln 962 GAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAG CTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC

985 STUI,

ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
1022 ACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
TGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGA

1069 DRA3,

ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeulleArgLeuLys
AGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAG

FIGURE 21 - Page 3

TCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC

- ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIle
 1142 CCCACCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATC
 GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG

 1150 NCOI.
- - 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
- ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGAAACCGGCGCATAACGGAC
- SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATA
 AGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTAT
 - 1369 NAEI,
- ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
 1382 CCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGAAGAGTGCTCTCAGCACTTA
 GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACCTTCTCACGAGAGTCGTGAAT

 1385 DRD1,
- ProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTC
 GGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAG
- LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGCCAGTCCCAATAGCGGGGACGACAGGTCTGGTTGACCGTT

 1502 PSTI, 1507 TTH31,
- - AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla GCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCT CGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGA
 - 1643 BSTE2, 1677 ALWN1 PVU2,
 - AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrp
 1682 GCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGG
 CGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACC

FIGURE 21 - Page 4

. 1794 ESP1,

GİYALAALAILEGİYSERVALĞİYLEUĞİYLYSVALLEUİLEASPILELEUALAĞİYTYR 1802 GGCGCCGCCATCGGCAĞTGTTĞĞACTĞĞĞĞAAĞĞTCCTCATAĞACATCCTTĞCAĞĞĞTAT CCGCĞĞCĞĞTAĞCCĞTCACAACCTĞACCCCTTCCAĞĞAĞTATCTĞTAĞĞAACĞTCCCATA

1802 KAS1 NARI.

GlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSer
1862 GGCGCGGGCCTGGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCC
CCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGG

1878 SACI, 1899 BSPH1.

ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
1922 ACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGC
TGCCTCCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGCCTCGGGAGCATCAGCCG

1928 TTH3I.

ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
1982 GTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGG
CACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACC

2004 NAEI, 2017 SMAI XMAI,

MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
2042 ATGAACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTG
TACTTGGCCGACTATCGGAAGCGGAGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCAC

2067 SMAI XMAI, 2093 DRA3,

ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
CCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAG
GGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC

2115 PVU2, 2159 ALWN1,

LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
2162 CTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
GAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAAG

2164 MST2, 2220 ECON1,

TrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
2222 TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
ACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGAT

2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 21 - Page 5

LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
2342 AAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTCGCCACTGTGGAGCTGAGATC
TTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG

- ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC
 - 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,
- TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 TGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGC

2480 ASE1, 2497 APAI,

ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
GGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC

5

2553 PSTI,

ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
2582 GTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAG
CACCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGCACGGTC

2594 DRA3,

- ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro
 2642 GTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCC
 CAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGG
- ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCG
 GGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC

2757 HGIE2.

ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
2762 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTC
CATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACTGCAGGTACGAG

2809 AAT2,

ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
2822 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGCCGAAGGTTGGCGAGGGGATCACCC
TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGG

2850 EAG1 XMA3,

ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGC
GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG

2889 BALI, 2903 NHEI,

FIGURE 21 - Page 6

- ThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGin
 2942 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCCGAGTATCTCCGGTTGGAGGATACCTCCGTC
 - _ 2966 ESP1, 2969 SACI,
- GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 3002 GAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGG
- PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGGGGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTCCTCGCCCCTCTAGAGGCATGGGCGTCTTTAGGAC
 - 3096 BGL2,
- ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro
 3122 CGGAAGTCTCGGAGGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCC
 GCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGCCAAACCCGCGCCGGCCTGATATTGGGG
 - 3143 ALWN1, 3164 EAG1 XMA3,
- ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 - 3217 HGIE2, 3229 NCOI,
- LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAG
- ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer
 3302 ACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCC
 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 - 3332 SACI, 3346 HIND3,
- CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGluGluPro
 3422 TGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGA
 - 3437 EAM11051,
- GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTC
 - 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
- AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCC
 CTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGG

FIGURE 21 - Page 7

3589 DRA3, 3600 SAC2,

- AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602. GCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAAT
 CGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGTTTA
 - 3611 ALWN1, 3655 PFLM1,
- LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 3662 TTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGAC
 AACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTG
 - 3681 DRA3,
- ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCG
 TCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGC
- SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 TCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCACAC
 AGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTG
 - 3816 HIND3,
- SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 TCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC
 AGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGG
 - 3875 AAT2, 3890 BGLI,
- ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG
- ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 3962 ACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTC
- ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTG
 GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAAC
- TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC
 ATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATG
- SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 4142 TCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATG
 AGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTAC
 - 4160 ECORI,
- GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC

FIGURE 21 - Page 8

4229 DRD1, 4236 ALWN1,

GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGG

4301 BGLI, 4308 BALI,

LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
4322 CTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGC
GAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCG

4345 APAI,

- TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCGCGGGGGGGGGGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACG
- TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 4442 TAÇATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCCGTCGGACAGCTCCGAGGTCCTGACGTGGTACGAGCAC

4452 SMAI XMAI,

CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGGTCCAGGAGGACGCGGCGAGC
ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCG

4508 DRD1, 4511 TTH3I,

- LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGTGTT
- ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG

4637 SACI, ~~~

GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
4682 GGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGAGCT
CCGCGACCTTTCTCCCAGATGATGGAGTGGCACTGGGATGTTGGGGGGAGCGCTCTCGA

4731 NRUI.

- AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 4742 GCGTGGGAGACAGCAGACACCTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTT
 CGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAA
- AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu

FIGURE 21 - Page

4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
TCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTT

4893 BGL2.

ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCAC
GGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG

4954 NCOI,

SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCG
TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGC

5015 SPHI, 5035 KPNI,

ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
5042 CCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTTCTGGCCAGAGGA
GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT

5064 APAI, 5091 BALI,

GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAA
CCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTT

5113 NDEI,

5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,

SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCGGTGGATCTGGTTTTGC
TCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGCCGACCTAGACCAAAACG

5240 DRA3,

LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
5282 CTACTCCTGCTGCAGGGGTAGGCATCTACCTCCCCCAACCGAATGAGCACGAAT
GATGAGGACGACGACGTCCCATCCGTAGATGAGGAGGGGGTTGGCTTACTCGTGCTTA

5295 PSTI.

ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
5342 CCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGGCGGCGCGGGACGTCAAGTTC
GGATTTGGAGTTCTTTCTGGTTTGCATTGTGGTTGGCCGCCGGCGTCCTGCAGTTCAAG

- 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMAI XMAI,

ProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
5402 CCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTG
GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAAC

FIGURE 21 - Page 10

5449 APAI,

- GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
 5462 GGTGTGCGCGGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGCTGCTCTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGA
 - 5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,
- IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 5522 ATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCC
 TAGGGGTTCCGAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGG
 - 5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,
- LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 5582 CTCTATGGCAATGAGGGCTGCGGGTGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGG
 GAGATACCGTTACTCCCGACGCCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCC
- ProSerTrpGlyProThrAspProArgArgArgAerArgAsnLeuGlyLysVallleAsp
 CCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGAT
 GGATCGACCCCGGGGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCCATTCCAGTAGCTA
- ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValOC AM
 5702 ACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCTAATAGTCGAC
 TGGGAATGCACGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGATTATCAGCTG
 - 5724 HGIE2, 5755 SALI,

. . . .

5650 APAI, 5696 CLAI,

FIGURE 22 - Page 1

- MetAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsn 2 AGCTTACAAAACAAAATGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAAC TCGAATGTTTTGTTTTACCGACGTATACGTCGAGTCCCGATATTCCACGATCATGAGTTG
 - 1 HIND3, 24 NDEI, 52 SCAI,
- ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAsp 62 CCCTCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGAT GGGAGACAACGACGTTGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTA
 - 116 CLAI,
- ProAsnileArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSerThr
 122 CCTAACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACC
 GGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTCGGGGTAGTGCATGAGGTGG
- TyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCys
 182 TACGGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGGCGCTTATGACATAATAATTTGT
 ATGCCGTTCAAGGAACGGCTGCCGCCCACGAGCCCCCCGCGAATACTGTATTATTAAACA
- AspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGln
 242 GACGAGTGCCACTCCACGGATGCCACTCTTGGGCATTGGCACTGTCCTTGACCAA
 CTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTGACAGGAACTGGTT
- AlaGluthrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySerVal
 302 GCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCCCTCCGGGCTCCGTC
 CGTCTCTGACGCCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAG
 303 ALWN1,
- ThrVaiProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPhe
 362 ACTGTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTT
 TGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAA
- TyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHis
 422 TACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTTCTGTCAT
 ATGCCGTTCCGATAGGGGGAGCTTCATTAGTTCCCCCCCTCTGTAGAGTAGAAGACAGTA

FIGURE 22 - Page 2

- SerLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaVal
 TCAAAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTG
 AGTTTCTTCTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCAC
- AlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValVal
 542 GCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTG
 CGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCCGCTACAACAGCAGCAC

550 SAC2, 560 DRD1,

AlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVallleAspCysAsn
GCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAAT
CGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACGTTA

615 BSPH1,

- ThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrlleGluThrlle
 662 ACGTGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTTCACCATTGAGACAATC
 TGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGAAGTGGTAACTCTGTTAG
- ThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLys
 ACGCTCCCCCAAGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAG
 TGCGAGGGGGTTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCCTGACCGTCCCCTTC
- ProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerSer
 782 CCAGGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCC
 GGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGG

816 BGLI, 833 DRD1,

881 SACI,

- ThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeu
 902 ACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTT
 TGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCCTGGTAGAA
 - 931 SMAI XMAI,
- GluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGln
 GAATTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAG
 CTTAAAACCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTC

985 STUI,

ThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAla
1022 ACAAAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCT
TGTTTCGTCTCACCCCTCTTGGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGA

1069 DRA3,

ArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLys
AGGGCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAG

FIGURE 22 - Page 3

TCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACAAACTAAGCGGAGTTC

ProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIie
1142 CCCACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATC
GGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAG

1150 NCOI,

- - 1230 BSPH1, 1234 DRD1, 1237 AVA3, 1245 EAG1 XMA3, 1250 DRD1,
- ValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeu
 1262 GTCACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTG
 CAGTGCTCGTGGACCCACGAGCAACCGCCGCAGGACCGACGAAACCGGCGCATAACGGAC
- SerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIle
 1322 TCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATA
 AGTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTTCGGCCGTTAGTAT

1369 NAEI,

ProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeu
1382 CCTGACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTA
GGACTGTCCCTTCAGGAGATGGCTCTCAAGCTACCTTCTCACGAGAGTCGTGAAT

1385 DRD1,

- ProTyrileGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu
 1442 CCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTC
 GGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAG
- LeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGln
 1502 CTGCAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGCCAGACCAACTGGCAA
 GACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTT
 ^ ^

1502 PSTI, 1507 TTH3I,

LysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeu
1562 AAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTG
TTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATGAAC

1565 XHOI, 1586 NDEI,

AlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAla GCGGGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCT CGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGTCGA

1643 BSTE2, 1677 ALWN1 PVU2,

AlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleL uGlyGlyTrp
1682 GCTGTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGG
CGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCCCACC

FIGURE 22 - Page 4

ValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAla 1742 GTGGCTGCCCAGCTCGCCCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCT CACCGACGGTCGAGCGGCGGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGA

~ 1794 ESP1,

GlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyr

GGCGCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTAT

CCGCGGGGGTAGCCGTCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATA

1802 KAS1 NARI,

1878 SACI, 1899 BSPH1,

ThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGly
1922 ACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGC
TGCCTCCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGCCCTCGGGAGCATCAGCCG

1928 TTH3I,

ValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrp
1982 GTGGTCTGTGCAGCAATACTGCGCCGGCCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGG
CACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTCACC

2004 NAEI, 2017 SMAI XMAI,

MetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrVal
2042 ATGAACCGGCTGATAGCCTTCGCCTCCGGGGGAACCATGTTTCCCCCACGCACTACGTG
TACTTGGCCGACTATCGGAAGCGGAGGCCCCCTTGGTACAAAGGGGGTGCGTGATGCAC

2067 SMAI XMAI, 2093 DRA3,

ProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGln
CCGGAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAG
GGCCTCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGTC

2115 PVU2, 2159 ALWN1,

LeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySer
2162 CTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
GAGGACTCCGCTGACGTGGTCACCTATTCGAGCCTCACATGGTGAGGTACGAGGCCAAGG

2164 MST2, 2220 ECON1,

TrpLeuArgAsplleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeu
TGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTA
ACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGAT

2285 ESP1, 2300 PVU2, 2310 BAMHI,

FIGURE 22 - Page 5

- LysGlyValTrpArgGlyAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIle
 2342 AAGGGGGTCTGGCGAGGGGACGGCATCATGCACACTCGCCACTGTGGAGCTGAGATC
 TTCCCCCAGACCGCTCCCCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAG
- ThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMet
 ACTGGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATG
 TGACCTGTACAGTTTTTGCCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTGTAC
 - 2425 BSAB1, 2441 AVR2, 2448 SSE83871, 2449 PSTI,
- TrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAla
 2462 TGGAGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCG
 ACCTCACCCTGGAAGGGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGGAAGGACGC
 - 2480 ASE1, 2497 APAI,
- ProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGln
 CCGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCAG
 GGCTTGATGTGCAAGCGCGATACCTCCCACAGACGTCTCCTTATGCACCTCTATTCCGTC
 - 2553 PSTI,
- ValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGln
 2582 GTGGGGGACTTCCACTACGTGACGGTATGACTACTGACAATCTTAAATGCCCGTGCCAG
 CACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAATTTACGGGCACGGTC
 - 2594 DRA3,
- ValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaPro 2642 GTCCCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCC CAGGGTAGCGGGCTTAAAAAGTGTCTTAACCTGCCCCACGCGGATGTATCCAAACGCGGG
- ProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrPro
 CCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCG
 GGGACGTTCGGGAACGACGCCCTCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGC
 - 2757 HGIE2,
- ValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeu
 2762 GTAGGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTC
 CATCCCAGCGTTAATGGAACGCTCGGGCTTGCCCTGCACCGGCACAACTGCAGGTACGAG
 - 2809 AAT2,
- ThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerPro
 2822 ACTGATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCC
 TGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCCCGCTTCCAACCGCTCCCCTAGTGGG
 - 2850 EAG1 XMA3,
- ProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCys
 CCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGC
 GGGAGACACCGGTCGAGGAGCCGATCGGTCGATAGGCGAGGTAGAGAGTTCCGTTGAACG
 - 2889 BALI, 2903 NHEI,

FIGURE 22 - Page 6

- ThrAlaAsnHisAspSerProAspAlaGluL uIleGluAlaAsnLeuLeuTrpArgGln
 ACCGCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAG
 TGGCGATTGGTACTGAGGGGACTACGACTCCGATTCTCCGGTTGGAGGATACCTCCGTC
 - 2966 ESP1, 2969 SACI,

3096 BGL2,

- GluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSer
 GAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCC
 CTCTACCCGCCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTTCACCACTAAGACCTGAGG
- PheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeu
 TTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTG
 AAGCTAGGCGAACACCGCCTCCTCCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGAC
- ArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnPro
 3122 CGGAAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCC
 GCCTTCAGAGCCTCTAAGCGGGTCCGGGACGGCCAAACCCGCGCCGGCCTGATATTGGGG
 - 3143 ALWN1, 3164 EAG1 XMA3,
- ProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCysPro
 3182 CCGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGCCCG
 GGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACGGGC
 - 3217 HGIE2, 3229 NCOI,
- LeuProProProLysSerProProValProProProArgLysLysArgThrValValLeu
 3242 CTTCCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTC
 GAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCCTGCCACCAGGAG
- ThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSer

 ACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCC

 TGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGG
 - 3332 SACI, 3346 HIND3,
- CysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluPro
 3422 TGCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCTGGAGGGGGAGCCT
 ACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGA
 - 3437 EAM11051,
- GlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGlu
 3482 GGGGATCCGGATCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCGGAG
 CCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCCTC
 - 3484 BAMHI, 3485 BSAB1, 3487 BSPE1,
- AspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAla
 3542 GATGTCGTGTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGCGCC
 CTACAGCACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCACGCGG

FIGURE 22 - Page 7

3589 DRA3, 3600 SAC2, -

- AlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsn
 3602 GCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCACAAT
 CGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCGTTGAGCAACGATGCAGTGGTGTTA
 - 3611 ALWN1, 3655 PFLM1,
- LeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAsp
 TTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTTGAC
 AACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTTCAGTGTAAACTG
 - 3681 DRA3,
- ArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAla
 3722 AGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGGCG
 TCTGACGTTCAAGACCTGTCGGTAATGGTCCTGCATGAGTTCCTCCAATTTCGTCGCCGC
- SerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHis
 TCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCACAC
 AGTTTTCACTTCCGATTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGGTGTG
 - 3816 HIND3,
- SerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAla
 TCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAGGCC
 AGTCGGTTTAGGTTCAAACCAATACCCCGTTTTCTGCAGGCAACGGTACGGTCTTTCCGG
 - 3875 AAT2, 3890 BGLI,
- ValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIleAsp
 3902 GTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATAGAC
 CATTGGGTGTAGTTGAGGCACACCTTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTG
- ThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArgLys
 3962 ACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGTAAG
 TGATGGTAGTACCGATTCTTGCTCCAAAAGACGCAAGTCGGACTCTTCCCCCCAGCATTC
- ProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAlaLeu
 4022 CCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCTTTG
 GGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAAC
- TyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyr
 4082 TACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATAC
 ATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCCTTCGAGGATGCCTAAGGTTATG
- SerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMet
 4142 TCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCAATG
 AGTGGTCCTGTCGCCCAACTTAAGGAGCACGTTCGCACCTTCAGGTTCTTTTGGGGTTAC
 - 4160 ECORI,
- GlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArgThr
 4202 GGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACG
 CCCAAGAGCATACTATGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGC

FIGURE 22 - Page 8

4229 DRD1, 4236 ALWN1,

GluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSer
4262 GAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCC
CTCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTCGGGCGCACCGGTAGTTCAGG

4301 BGLI, 4308 BALI,

LeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGly
4322 CTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGC
GAGTGGCTCTCCGAAATACAACCCCCGGGAGAATGGTTAAGTTCCCCCCTCTTGACGCCG

4345 APAI,

- TyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCys
 4382 TATCGCAGGTGCCGCGGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACTTGC
 ATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACG
- TyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuVal
 4442 TACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTCGTG
 ATGTAGTTCCGGGCCCGTCGGACAGCTCCGGCGTCCCGAGGTCCTGACGTGGTACGAGCAC

4452 SMAI XMAI,

CysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSer
4502 TGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGGTCCAGGAGGACGCGGCGAGC
ACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCCTCCTGCGCCGCTCG

4508 DRD1, 4511 TTH3I,

- LeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProProGln
 4562 CTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCTGGGGACCCCCACAA
 GACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGGTGTT
- ProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHisAsp
 4622 CCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCACGAC
 GGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGCTG

4637 SACI,

GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
4682 GGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCCTACAACCCCCCTCGCGAGAGCT
CCGCGACCTTTCTCCCAGATGATGGGGGCACTGGGATGTTGGGGGGAGCGCTCTCGA

4731 NRUI,

- AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 4742 GCGTGGGAGACAGCAGACACCTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTTT
 CGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAA
- AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 4802 GCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCC
 CGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGG

4806 PFLM1, 4807 DRA3,

ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu

rIGURE 22 - Page 9

- 4862 AGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATAGAA
 TCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTT
 - 4893 BGL2,
- ProLeuAspLeuProProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHis
 4922 CCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCAC
 GGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTG
 - 4954 NCOI,
- SerTyrSerProGlyGluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValPro
 4982 AGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTACCG
 TCAATGAGAGGTCCACTTTAGTTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGC
 - 5015 SPHI, 5035 KPNI,
- ProLeuArgAlaTrpArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGly
 5042 CCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGAGGA
 GGGAACGCTCGAACCTCTGTGGCCCGGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCT
 - 5064 APAI, 5091 BALI,
- GlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys
 5102 GGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAA
 CCGTCCCGACGGTATACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTTCGAGTTT
 - 5113 NDEI,
- - 5174 NOTI, 5175 EAG1 XMA3, 5182 BALI, 5186 PVU2,
- SerGlyGlyAspIleTyrHisSerValSerHisAlaArgProArgTrpIleTrpPheCys
 5222 AGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGATCTGGTTTTGC
 TCGCCCCCTCTGTAAATAGTGTCGCACAGAGTACGGGCCGGGGCGACCTAGACCAAAACG
 - 5240 DRA3,
- LeuLeuLeuAlaAlaGlyValGlyIleTyrLeuLeuProAsnArgMetSerThrAsn
 CTACTCCTGCTGCTGCAGGGGTAGGCATCTACCTCCCCAACCGAATGAGCACGAAT
 GATGAGGACGAACGACGTCCCCATCCGTAGATGGAGGAGGGGTTGGCTTACTCGTGCTTA
 - 5295 PSTI,
 - ProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGlnAspValLysPhe
 5342 CCTAAACCTCAAAGAAAGACCAAACGTAACACCAACCGGCGGCCGCAGGACGTCAAGTTC
 GGATTTGGAGTTCTTTCTGGTTTGCATTGTGGTTGGCCGCCGGCGTCCTGCAGTTCAAG
 - 5380 NOTI, 5381 EAG1 XMA3, 5390 AAT2, 5401 SMAI XMAI,
 - ProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeu
 5402 CCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTAGATTG
 GGCCCACCGCCAGTCTAGCAACCACCTCAAATGAACAACGGCGCTCCCCGGGATCTAAC

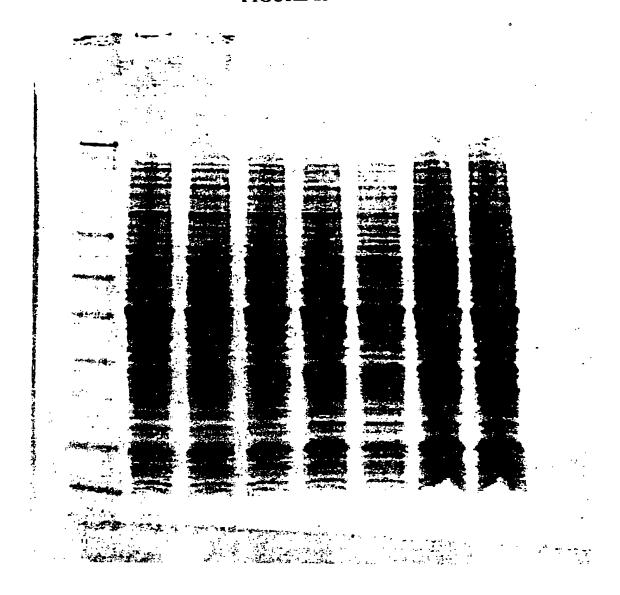
FIGURE 22 - Page 10

5449 APAI,

- GlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnPro
 5462 GGTGTGCGCGACGAGAAAGACTTCCGAGCGTCGCAACCTCGAGGTAGACGTCAGCCT
 CCACACGCGCGCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGA
 - 5467 BSSH2, 5478 XMNI, 5502 XHOI, 5511 AAT2,
- IleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpPro
 5522 ATCCCCAAGGCTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCC
 TAGGGGTTCCGAGCAGCCGGGCTCCCGTCCTGGACCCGAGTCGGGCCCATGGGAACCGGG
 - 5548 ALWN1, 5558 ESP1, 5564 SMAI XMAI, 5568 KPNI,
- LeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArg
 5582 CTCTATGGCAATGAGGGCTGGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGG
 GAGATACCGTTACTCCCGACGCCCACCGGCCCTACCGAGGAGAGAGCC
- ProSerTrpGlyProThrAspProArgArgArgSerArgAsnLeuGlyLysVallleAsp
 5642 CCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGAT
 GGATCGACCCCGGGGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTA
 ^
 - 5650 APAI, 5696 CLAI,
- ThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeu
 5702 ACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCCCTCTT
 TGGGAATGCACGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGCCGCGGGGAGAA
 - 5724 HGIE2, 5750 KAS1 NARI, 5756 ECON1,
- GlyGlyAlaAlaArgAlaOC AM
 5762 GGAGGCGCTGCCAGGGCCTAATAGTCGAC
 CCTCCGCGACGGTCCCGGATTATCAGCTG

5785 SALI,

FIGURE 23



SEQUENCE LISTING

```
<110> CHIRON CORPORATION et al.
<120> NOVEL HCV NON-STRUCTURAL POLYPEPTIDE
<130> PP01617.003
<140>
<141>
<160> 19
<170> PatentIn Ver. 2.0
<210> 1
<211> 9620
<212> DNA
<213> Artificial Sequence
<220>
<221> CDS
<222> (1990)..(7302)
<223> Description of Artificial Sequence: Hepatitis C pns345
<400>1
cgcgcgtttc ggtgatgacg gtgaaaacct ctgacacatg cagctcccgg agacggtcac 60
agettgtetg taageggatg eegggageag acaageeegt eagggegegt cagegggtgt 120
tggcgggtgt cggggctggc ttaactatgc ggcatcagag cagattgtac tgagagtgca 180
ccatatgaag ctttttgcaa aagcctaggc ctccaaaaaa gcctcctcac tacttctgga 240
atageteaga ggeegaggeg geeteggeet etgeataaat aaaaaaaatt agteageeat 300
ggggcggaga atgggcggaa ctgggcgggg agggaattat tggctattgg ccattgcata 360
cgttgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420
gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 480
gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 540
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 600
ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtac 660
atcaagtgta tcatatgcca agtccgccc ctattgacgt caatgacggt aaatggcccg 720
cetggcatta tgcccagtac atgacettac gggaetttec tacttggcag tacatetacg 780
tattagtcat cgctattacc atggtgatgc ggttttggca gtacaccaat gggcgtggat 840
agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 900
```

tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc ccgttgacqc 960 aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctcgt ttagtgaacc 1020 gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga caccgggacc 1080 gatecageet eegeggeegg gaacggtgea ttggaacgeg gatteceegt gecaagagtg 1140 acgtaagtac cgcctataga ctctataggc acaccccttt ggctcttatg catgctatac 1200 tgtttttggc ttggggccta tacacccccg ctccttatgc tataggtgat ggtatagctt 1260 agcetatagg tgtgggttat tgaccattat tgaccactee cetattggtg acgatacttt 1320 ccattactaa tocataacat ggctctttgc cacaactatc tctattggct atatgccaat 1380 actetyteet teagagactg acaeggacte tytattttta caggatyggg tecatttatt 1440 atttacaaat tcacatatac aacaacgccg tcccccgtgc ccgcagtttt tattaaacat 1500 agogtgggat ctccgacatc tcgggtacgt gttccggaca tgggctcttc tccggtagcg 1560 geggagette cacateegag ceetggteee ateegteeag eggeteatgg tegeteggea 1620 gctccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800 agttgttgta ttctgataag agtcagaggt aactcccgtt gcggtgctgt taacggtgga 1860 gggcaqtgta qtctqaqcag tactcqttqc tqccqcqcqc qccaccaqac ataataqctq 1920 acagactaac agactgttcc tttccatggg tcttttctgc agtcaccgtc gtcgacctaa 1980 gaattcacc atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc 2031 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac atg tcc aag 2079 Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys gct cat ggg atc gat cct aac atc agg acc ggg gtg aga aca att acc 2127 Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc gac 2175 Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt gac gag tgc 2223 Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys 70 cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act gtc ctt gac His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp 85

					999 Gly 100									2319
					act Thr									2367
					gag Glu									2415
_	_		_		gly ggg	_					_		_	 2463
					gcc Ala									2511
					ggt Gly 180									2559
					gca Ala									2607
					ata Ile									2655
-		_			cct Pro							_		2703
					cgc Arg									2751
aag Lys 255					aga Arg 260						Glu			2799
					gtc Val									2847
				_	ccc Pro	_				_		_		2895
_			_		ctt Leu		_	-	_	-		-		 2943
					ggc Gly					_	_			2991

 $\varphi_{a}(Q)$

_		-	_	agt Ser		_					_	_				3039
_			_	gct Ala 355		_		_				_		_	•	3087
_		_	_	ttg Leu		-		-								3135
				aga Arg												3183
				aaa Lys												3231
				acc Thr												3279
				ctg Leu 435												3327
				aag Lys												3375
				gag Glu	_	_		-		_			_			3423
				atg Met												3471
				gcg Ala		Arg										3519
cag Gln	acc Thr	aac Asn	tgg Trp	caa Gln 515	aaa Lys	ctc Leu	gag Glu	acc Thr	ttc Phe 520	tgg Trp	gcg Ala	aag Lys	cat His	atg Met 525	tgg Trp	3567
			_	gly aaa				_			_		_	_		3615
				att Ile												3663
agc Ser	cca Pro 560	cta Leu	acc Thr	act Thr	agc Ser	caa Gln 565	acc Thr	ctc Leu	ctc Leu	ttc Phe	aac Asn 570	ata Ile	ttg Leu	ggg Gly	Gly ggg	3711

						gcc Ala										3759
						gcc Ala										3807
						gca Ala										3855
						atg Met										3903
						gcc Ala 645										3951
						ata Ile										3999
						aac Asn										4047
aac Asn	cat His	gtt Val	tcc Ser 690	ccc Pro	acg Thr	cac His	tac Tyr	gtg Val 695	ccg Pro	gag Glu	agc Ser	gat Asp	gca Ala 700	gct Ala	gcc Ala	4095
						agc Ser										4143
						agc Ser 725										4191
					Ile	tgg Trp	Asp	${\tt Trp}$	Ile	Сув	Glu					4239
						gct Ala										4287
ccc Pro	ttt Phe	gtg Val	tcc Ser 770	tgc Cys	cag Gln	cgc Arg	Gly 999	tat Tyr 775	aag Lys	Gly ggg	gtc Val	tgg Trp	cga Arg 780	Gly aaa	gac Asp	4335
ggc Gly	atc Ile	atg Met 785	cac His	act Thr	cgc Arg	tgc Cys	cac His 790	tgt Cys	gga Gly	gct Ala	gag Glu	atc Ile 795	act Thr	gga Gly	cat His	4383
gtc Val	aaa Lys 800	aac Asn	999 Gly	acg Thr	atg Met	agg Arg 805	atc Ile	gtc Val	ggt Gly	cct Pro	agg Arg 810	acc Thr	tgc Cys	agg Arg	aac Asn	4431

atg tgg agt ggg Met Trp Ser Gly 815					
acc ccc ctt cct Thr Pro Leu Pro		_			
gca gag gaa tac Ala Glu Glu Tyr 850					
acg ggt atg act Thr Gly Met Thr 865	Thr Asp Asn	_		_	_
ccc gaa ttt ttc Pro Glu Phe Phe 880					
ccc ccc tgc aag Pro Pro Cys Lys 895					
ctc cac gaa tac Leu His Glu Tyr					
gac gtg gcc gtg Asp Val Ala Val 930	Leu Thr Ser				
gca gag gcg gcc Ala Glu Ala Ala 945	Gly Arg Arg				
gcc agc tcc tcg Ala Ser Ser Ser 960					
tgc acc gct aac Cys Thr Ala Asn 975	His Asp Ser		Glu Leu Ile		
ctc cta tgg agg Leu Leu Trp Arg					
gaa aac aaa gtg Glu Asn Lys Val 1010	Val Ile Leu	gac tcc ttc Asp Ser Phe 1015	Asp Pro Leu	gtg gcg Val Ala 1020	gag 5055 Glu
gag gac gag cgg Glu Asp Glu Arg 1025	Glu Ile Ser				
cgg aga ttc gcc Arg Arg Phe Ala 1040	cag gcc ctg Gln Ala Leu 1045	ccc gtt tgg Pro Val Trp	gcg cgg ccg Ala Arg Pro 1050	gac tat Asp Tyr	aac 5151 Asn

ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa cca cct gtg Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val 1055 1060 1065 1070	5199
gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct gtg cct ccg Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro 1075 1080 1085	5247
cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc cta tct act Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr 1090 1095 1100	5295
gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc tca act tcc Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser 1105 1110 1115	5343
ggc att acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct tct Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser 1120 1130	5391
ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc atg ccc ccc Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro 1135 1140 1145 1150	5439
ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg tca tgg tca Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser 1155 1160 1165	5487
acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc tgc tca atg Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met 1170 1180	5535
tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc gcg gaa gaa Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu 1185 1190 1195	5583
cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta cgt cac cac Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His 1200 1205 1210	5631
aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa agg cag aag Asn Leu Val Tyr Ser Thr Thr Ser Arg: Ser Ala Cys Gln Arg Gln Lys 1215 1220 1225 1230	5679
aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag gac Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp 1235 1240 1245	5727
gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag gct aac ttg Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu 1250 1260	5775
cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac tca gcc aaa Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys 1265 1270 1275	5823
tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat gcc aga aag Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys 1280 1285 1290	5871

gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac aat Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn 1295 1300 1305 1310	5919
gta aca cca ata gac act acc atc atg gct aag aac gag gtt ttc tgc Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys 1315 1320 1325	5967
gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg ttc Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe 1330 1335 1340	6015
ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg tac gac gtg Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val 1345 1350 1355	6063
gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac gga ttc caa Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln 1360 1365 1370	6111
tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg tgg aag tcc Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser 1375 1380 1385 1390	6159
aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc ttt gac tcc Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser 1395 1400 1405	6207
aca gtc act gag agc gac atc cgt acg gag gag gca atc tac caa tgt Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys 1410 1415 1420	6255
tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc ctc acc gag Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu 1425 1430 1435	6303
agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg gag aac tgc Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys 1440 1445 1450	6351
ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act agc tgt ggt Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly 1455 1460 1465 1470	6399
aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt cga gcc gca Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala 1475 1480 1485	6447
ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac tta gtc gtt Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val 1490 1495 1500	6495
atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc ctg aga gcc Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala 1505 1510 1515	6543
Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro 1520 1525 1530	6591

t., •

caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc tcc aac gtg Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val 1535 1540 1545 1550	6639
tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac ctc acc cgt Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg 1555 1560 1565	6687
gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca gca aga cac Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His 1570 1575 1580	6735
act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc aca Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr 1585 1590 1595	6783
ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctt ata Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile 1600 1605 1610	6831
gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc tac ggg gcc Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala 1615 1620 1625 1630	6879
tgc tac tcc ata gaa cca ctg gat cta cct cca atc att caa aga ctc Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu 1635 1640 1645	6927
cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa atc His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile 1650 1660	6975
aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg ccc ttg cga Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg 1665 1670 1675	7023
gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg gcc aga Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg 1680 1685 1690	7071
gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca gta Gly Gly Arg Ala Ala Ile Cys Gly: Lys Tyr Leu Phe Asn Trp Ala Val 1695 1700 1705 1710	7119
aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc cag ctg gac Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp 1715 1720 1725	7167
ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac att tat cac Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His 1730 1740	7215
age gtg tet cat gee egg eee ege tgg ate tgg ttt tge eta ete etg Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu 1745 1750 1755	7263

ggtaaacact ccggcctaaa aaaaaaaaaa aatctagaaa ggcgcgccaa gatatcaagg 7372 atccactacg cgttagagct cgctgatcag cctcgactgt gccttctagt tgccagccat 7432 ctgttgtttg cccctcccc gtgccttcct tgaccctgga aggtgccact cccactgtcc 7492 tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat tctattctgg 7552 ggggtggggt ggggcaggac agcaaggggg aggattggga agacaatagc aggcatgctg 7612 gggagetett eegetteete geteaetgae tegetgeget eggtegtteg getgeggega 7672 geggtateag eteaeteaaa ggeggtaata eggttateea eagaateagg ggataaegea 7732 ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg 7792 ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg acgctcaagt 7852 cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc 7912 etegtgeget eteetgttee gaccetgeeg ettaceggat acetgteege ettteteeet 7972 tegggaageg tggegettte teaatgetea egetgtaggt ateteagtte ggtgtaggte 8032 gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta 8092 teeggtaact ategtettga gteeaaceeg gtaagacaeg acttategee aetggeagea 8152 gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 8212 tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag 8272 ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt 8332 ageggtggtt tttttgtttg caageageag attacgegea gaaaaaaagg atetcaagaa 8392 gateetttga tetttetae ggggtetgae geteagtgga acgaaaacte acgttaaggg 8452 attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga 8512 agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta 8572 atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc 8632 cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg 8692 ataccgcgag acccacgete accggeteca gatttateag caataaacca gecageegga 8752 agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt 8812 tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt 8872 gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc 8932 caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc 8992 ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca 9052 gcactgcata attetettae tgteatgcea teegtaagat gettttetgt gaetggtgag 9112

tactcaacca agtcattctg agaatagtgt atgeggegac egagttgete ttgeceggeg 9172
tcaatacggg ataataccge gccacatage agaactttaa aagtgeteat cattggaaaa 9232
cgttettegg ggcgaaaact etcaaggate ttacegetgt tgagatecag ttegatgtaa 9292
cccactegtg cacceaactg atetteagea tetttaett teaceagegt ttetgggtga 9352
gcaaaaaacag gaaggcaaaa tgeegcaaaa aagggaataa gggegacaeg gaaatgttga 9412
atacteatae tetteettt teaatattat tgaagcattt ateagggtta ttgteteatg 9472
ageggataca tattgaatg tatttagaaa aataaacaaa taggggttee gegeacattt 9532
ccccgaaaag tgeeacetga egtetaagaa accattatta teatgacatt aacctataaa 9592
aataggegta teaegaggee etttegte

<210> 2

<211> 1771

<212> PRT

<213> Hepatitis C virus

<220>

<223> Description of Artificial Sequence: Hepatitis C pns345

<400> 2

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala

165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
245 250 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe 260 265 270

Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr 275 280 285

Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 290 295 300

Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 305 310 315 320

Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr 340 345 350

Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp 355 360 365

Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu 370 380

Leu Tyr Arg: Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro 385 390 395 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala 420 425 430

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu 435 440 445

Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 450 455 460

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 465 470 475 480

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 550 555 Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val 570 Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val 645 650 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala 665 Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val 695 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 710 715 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp 725 730 Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe 760 Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys

795

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815

- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845
- Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 855 860
- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120

Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135

- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455

- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1735 1740
- Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg

<210> 3 <211> 9620 <212> DNA <213> Artificial Sequence <220> <221> CDS <222> (1990)..(7302) <220> <223> Description of Artificial Sequence: pDeltaNS3NS5 egegegttte ggtgatgaeg gtgaaaacet etgacacatg cageteeegg agaeggteac 60 agettgtetg taageggatg cegggageag acaageeegt cagggegegt cagegggtgt 120 tggcgggtgt cggggctggc ttaactatgc ggcatcagag cagattgtac tgagagtqca 180 ccatatgaag ctttttgcaa aagcctaggc ctccaaaaaa gcctcctcac tacttctgga 240 atageteaga ggeegaggeg geeteggeet etgeataaat aaaaaaaatt agteageeat 300 ggggcggaga atgggcggaa ctgggcgggg agggaattat tggctattgg ccattgcata 360 egttgtatet atateataat atgtacattt atattggete atgtecaata tgacegecat 420 gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 480 gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 540 ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 600 ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtac 660 atcaagtgta tcatatgcca agtccgccc ctattgacgt caatgacggt aaatggcccg 720 cctggcatta tgcccagtac atgaccttac gggactttcc tacttggcag tacatctacg 780 tattagtcat cgctattacc atggtgatgc ggttttggca gtacaccaat gggcgtggat 840 ageggtttga ctcaegggga tttccaagtc tccaecccat tgaegtcaat gggagtttgt 900 tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc ccqttqacqc 960 aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctcgt ttagtgaacc 1020 gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga caccgggacc 1080 gatecageet eegeggeegg gaacggtgca ttggaacgeg gatteeeegt gecaagagtg 1140 acgtaagtac cgcctataga ctctataggc acaccccttt ggctcttatg catgctatac 1200 tgtttttggc ttggggccta tacacccccg ctccttatgc tataggtgat ggtatagctt 1260

agoctatagg tgtgggttat tgaccattat tgaccactcc cctattqqtg acqatacttt 1320 ccattactaa tccataacat ggctctttgc cacaactatc tctattggct atatgccaat 1380 actotyteet teagagacty acaeggacte tytattttta caqqatqqqq tecatttatt 1440 atttacaaat tcacatatac aacaacgccg tcccccgtgc ccgcagtttt tattaaacat 1500 agegtgggat ctccgacatc tcgggtacgt gttccggaca tgggctcttc tccggtagcg 1560 geggagette cacateegag ceetggteec ateegteeag eggeteatgg tegeteggea 1620 gctccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800 agttgttgta ttctgataag agtcagaggt aactcccgtt gcggtgctgt taacggtgga 1860 gggcagtgta gtctgagcag tactcgttgc tgccgcgcgc gccaccagac ataatagctg 1920 acagactaac agactgttcc tttccatggg tcttttctgc agtcaccgtc gtcgacctaa 1980 gaattcacc atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc 2031 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu aac ecc tet gtt get gea aca etg gge ttt ggt get tae atg tee aag 2079 Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys get cat ggg atc gat cet aac atc agg acc ggg gtg aga aca att acc 2127 Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc gac 2175 Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt gac gag tgc 2223 Gly Gly Cys Ser Gly: Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys cac tee acg gat gee aca tee ate ttg gge att gge act gte ett gae 2271 His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp 85 caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc acc gcc acc Gin Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr 100 105 cct ccg ggc tcc gtc act gtg ccc cat ccc aac atc gag gag gtt gct 2367 Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala 115 120 ctg tcc acc acc gga gag atc cct ttt tac ggc aag gct atc ccc ctc 2415 Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu 135

					ggg Gly								2463
					gcc Ala								2511
					ggt Gly 180								2559
_	-	-	-		gca Ala		_		_				2607
					ata Ile								2655
_		-		_	cct Pro				_		_		2703
	_	_	_		cgc Arg		_					 	2751
					aga Arg 260								2799
					gtc Val								2847
					ccc Pro								2895
_			_		ctt Leu	Val	_	Gln	Asp	His			2943
					ggc Gly								2991
					999 Gly 340								3039
					agg Arg								3087
					att Ile								3135

				aga Arg												3183
		_		aaa Lys			_		_	_	_	_	_	_		3231
				acc Thr												3279
_			_	ctg Leu 435				_		_					-	3327
_	_			aag Lys	_	_				-		_	-			3375
				gag Glu												3423
			_	atg Met		-		_		_	_	_	_			3471
				gcg Ala												3519
				caa Gln 515												3567
				Gly ggg												3615
			_	att Ile	_		_	_	_			_	_	_		3663
_				act Thr	_								_			3711
				cag Gln												3759
				gct Ala 595												3807
				atc Ile												3855

		gca Ala 625														3903
_	_	aat Asn												_	_	3951
	_	gtc Val	_	_					_		_					3999
		gtg Val						-								4047
		gtt Val			_				_		_	_	_	_	_	4095
- ,	_	act Thr 705	_			_	_			_		_		_		4143
_	_	cac His	_			_	_		_				_			4191
		cta Leu		_			_			_			_	_	_	4239
	_	acc Thr				_	_		_		_	_				4287
		gtg Val														4335
		atg Met 785			_	_		_		_						4383
_		aac Asn		_	_			_					_			4431
		agt Ser														4479
		ctt Leu			_			_								4527
-		gaa Glu						_			-					4575

acg ggt atg act Thr Gly Met Thr 865	act gac aat Thr Asp Asn	ctt aaa tgc Leu Lys Cys 870	ccg tgc cag g Pro Cys Gln V 875	tc cca tcg al Pro Ser	4623
ccc gaa ttt ttc Pro Glu Phe Phe 880	aca gaa ttg Thr Glu Leu 885	gac ggg gtg Asp Gly Val	cgc cta cat a Arg Leu His A 890	gg ttt gcg rg Phe Ala	4671
ccc ccc tgc aag Pro Pro Cys Lys 895	ccc ttg ctg Pro Leu Leu 900	cgg gag gag Arg Glu Glu	gta tca ttc a Val Ser Phe A 905	ga gta gga rg Val Gly 910	4719
ctc cac gaa tac Leu His Glu Tyr					4767
gac gtg gcc gtg Asp Val Ala Val 930			Asp Pro Ser H		4815
gca gag gcg gcc Ala Glu Ala Ala 945					4863
gcc agc tcc tcg Ala Ser Ser Ser 960					4911
tgc acc gct aac Cys Thr Ala Asn 975	cat gac tcc His Asp Ser 980	cct gat gct Pro Asp Ala	gag ctc ata g Glu Leu Ile G 985	ag gcc aac lu Ala Asn 990	4959
ctc cta tgg agg Leu Leu Trp Arg	cag gag atg Gln Glu Met 995	ggc ggc aac Gly Gly Asn 1000	atc acc agg g Ile Thr Arg V	tt gag tca al Glu Ser 1005	5007
gaa aac aaa gtg Glu Asn Lys Val 1010				al Ala Glu	5055
gag gac gag cgg Glu Asp Glu Arg 1025	Glu Ile Ser				5103
cgg aga ttc gcc Arg Arg Phe Ala 1040	cag gcc ctg Gln Ala Leu 1045	ccc gtt tgg Pro Val Trp	gcg cgg ccg g Ala Arg Pro A 1050	ac tat aac sp Tyr Asn	5151
ccc ccg cta gtg Pro Pro Leu Val 1055	gag acg tgg Glu Thr Trp 1060	Lys Lys Pro	gac tac gaa c Asp Tyr Glu P 1065	ca cct gtg ro Pro Val 1070	5199
gtc cat ggc tgc Val His Gly Cys 1	ccg ctt cca Pro Leu Pro 1075	cct cca aag Pro Pro Lys 1080	tcc cct cct g Ser Pro Pro V	tg cct ccg al Pro Pro 1085	5247
cct cgg aag aag Pro Arg Lys Lys 1090	cgg acg gtg Arg Thr Val	gtc ctc act Val Leu Thr 1095	gaa tca acc c Glu Ser Thr L 11	eu Ser Thr	5295

gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc tca act tcc Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser 1105 1110 1115	5343
ggc att acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct tct Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser 1120 1125 1130	5391
ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc atg ccc ccc Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro 1135 1140 1145 1150	5439
ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg tca tgg tca Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser 1155 1160 1165	5487
acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc tgc tca atg Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met 1170 1175 1180	5535
tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc gcg gaa gaa Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu 1185 1190 1195	5583
cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta cgt cac cac Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His 1200 1205 1210	5631
aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa agg cag aag Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys 1215 1220 1225 1230	5679
aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag gac Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp 1235 1240 1245	5727
gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag gct aac ttg Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu 1250 1255 1260	5775
cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac tca gcc aaa Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys 1265 1270 1275	5823 ***
tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat gcc aga aag Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys 1280 1285 1290	5871
gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac aat Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn 1295 1300 1305 1310	5919
gta aca cca ata gac act acc atc atg gct aag aac gag gtt ttc tgc Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys 1315 1320 1325	5967
gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg ttc Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe 1330 1335 1340	6015

\$3m

		atg gct ttg tac gac gtg Met Ala Leu Tyr Asp Val 1355	6063
-		agc tcc tac gga ttc caa Ser Ser Tyr Gly Phe Gln 1370	6111
	g Val Glu Phe Leu	gtg caa gcg tgg aag tcc Val Gln Ala Trp Lys Ser 1390	6159
		acc cgc tgc ttt gac tcc Thr Arg Cys Phe Asp Ser 1405	6207
		gag gca atc tac caa tgt Glu Ala Ile Tyr Gln Cys 1420	6255
		atc aag tcc ctc acc gag Ile Lys Ser Leu Thr Glu 1435	6303
		tca agg ggg gag aac tgc Ser Arg Gly Glu Asn Cys 1450	6351
	g Ala Ser Gly Val	ctg aca act agc tgt ggt Leu Thr Thr Ser Cys Gly 1470	6399
		gca gcc tgt cga gcc gca Ala Ala Cys Arg Ala Ala 1485	6447
		ggc gac gac tta gtc gtt Gly Asp Asp Leu Val Val 1500	6495
atc tgt gaa agc gcg gg Ile Cys Glu Ser Ala Gl 1505	g gtc cag gag gac y Val Gln Glu Asp 1510	gcg gcg agc ctg aga gcc Ala Ala Ser Leu Arg Ala 1515	6543
		ccc cct ggg gac ccc cca Pro Pro Gly Asp Pro Pro 1530	6591
	u Glu Leu Ile Thr	tca tgc tcc tcc aac gtg Ser Cys Ser Ser Asn Val 1550	6639
		gtc tac tac ctc acc cgt Val Tyr Tyr Leu Thr Arg 1565	6687
gac cct aca acc ccc ct Asp Pro Thr Thr Pro Le 1570	c gcg aga gct gcg u Ala Arg Ala Ala 1575	tgg gag aca gca aga cac Trp Glu Thr Ala Arg His 1580	6735

act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc aca 6783 Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr 1585 1590 1595	3
ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctt ata 6833 Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile 1600 1605 1610	1
gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc tac ggg gcc 6879 Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala 1615 1620 1625 1630	€
tgc tac tcc ata gaa cca ctg gat cta cct cca atc att caa aga ctc 6927 Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu 1635 1640 1645	7
cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa atc 6975 His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile 1650 1655 1660	5
aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg ccc ttg cga 7023 Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg 1665 1670 1675	3
gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg gcc aga 7073 Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg 1680 1685 1690	1
gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca gta 7119 Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val 1695 1700 1705 1710	9
aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc cag ctg gac 7167 Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp 1715 1720 1725	7
ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac att tat cac 7219 Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His 1730 1735 1740	5
age gtg tet cat gee egg eee ege tgg ate tgg ttt tge eta ete etg 7260 Ser Val Ser His. Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu 1745 1750 1755	3
ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga tgaaggttgg 731: Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1760 1765 1770	2
ggtaaacact ccggcctaaa aaaaaaaaa aatctagaaa ggcgcgccaa gatatcaagg 737	2
atccactacg cgttagagct cgctgatcag cctcgactgt gccttctagt tgccagccat 743	2
ctgttgtttg cccctccccc gtgccttcct tgaccctgga aggtgccact cccactgtcc 749	2
tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat tctattctgg 755	2
ggggtggggt ggggcaggac agcaaggggg aggattggga agacaatagc aggcatgctg 761	2
gggagetett eegetteete geteactgae tegetgeget eggtegtteg getgeggega 767	2

gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca 7732 ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg 7792 ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg acgctcaagt 7852 cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc 7912 ctegtgeget etectgttee gaccetgeeg ettaceggat acetgteege ettteteeet 7972 tegggaageg tggegettte teaatgetea egetgtaggt ateteagtte ggtgtaggte 8032 gttegeteca agetgggetg tgtgcaegaa ceceegtte ageeegaeeg etgegeetta 8092 teeggtaaet ategtettga gteeaaceeg gtaagacaeg aettategee aetggeagea 8152 gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 8212 tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag 8272 ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt 8332 ageggtggtt tttttgtttg caageageag attacgegea gaaaaaaaagg ateteaagaa 8392 gatcetttga tetttetae ggggtetgae geteagtgga acgaaaacte acgttaaggg 8452 attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga 8512 agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta 8572 atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc 8632 cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg 8692 atacegegag acceaegete aceggeteca gatttateag caataaacca gecageegga 8752 agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt 8812 tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt 8872 gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc 8932 caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc 8992 ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca 9052 gcactgcata attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag 9112 tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg 9172 tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa 9232 cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa 9292 cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga 9352 gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga 9412 atactcatac tetteetttt teaatattat tgaageattt ateagggtta ttgteteatg 9472

agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt 9532 ccccgaaaag tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa 9592 aataggcgta tcacgaggcc ctttcgtc 9620

<210> 4

<211> 1771

<212> PRT

<213> Artificial Sequence

<400> 4

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 215 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro

PCT/US00/32326

WO 01/38360

245 250 255 Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 295 Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 310 Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val 410 Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 455 Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 470 475 Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr 505 Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 520 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn 535

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro

550

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val 695 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 710 His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp 725 730 Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 810 Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 840 Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 870 875

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895

- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100

1000

- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys 185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215

- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520

Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
1525 1530 1535

- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1735 1740
- Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala 745 1750 1755 1760
- Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1765 1770
- <210> 5
- <211> 4282
- <212> DNA
- <213> Artificial Sequence
- <220>
- <223> Description of Artificial Sequence: pCMVII
- <400> 5
- tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea 60

cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg 120 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180 accatatgaa getttttgca aaageetagg eeteeaaaaa ageeteetea etaettetgg 240 aatageteag aggeegagge ggeeteggee tetgeataaa taaaaaaaat tagteageea 300 tggggcggag aatgggcgga actgggcggg gagggaatta ttggctattg gccattgcat 360 acgttgtatc tatatcataa tatgtacatt tatattggct catgtccaat atgaccgcca 420 tgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat 480 agcecatata tggagtteeg egttacataa ettaeggtaa atggeeegee tggetgaeeg 540 cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata 600 gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta 660 catcaagtgt atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc 720 gcctggcatt atgcccagta catgacctta cgggactttc ctacttggca gtacatctac 780 gtattagtca tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga 840 tageggtttg acteaegggg attteeaagt etceaececa ttgaegteaa tgggagtttg 900 ttttggcacc aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg 960 caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac 1020 cgtcagatcg cctggagacg ccatccacgc tgttttgacc tccatagaag acaccgggac 1080 cgatccagcc tccgcggccg ggaacggtgc attggaacgc ggattccccg tgccaagagt 1140 gacgtaagta ccgcctatag actctatagg cacacccctt tggctcttat gcatgctata 1200 ctgtttttgg cttggggcct atacaccccc gcttccttat gctataggtg atggtatagc 1260 ttagcctata ggtgtgggtt attgaccatt attgaccact cccctattgg tgacgatact 1320 ttccattact aatccataac atggctcttt gccacaacta tctctattgg ctatatgcca 1380 atactetgte etteagagae tgacaeggae tetgtatttt tacaggatgg ggteceattt 1440 attatttaca aattcacata tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa 1500 catagogtgg gatctccacg cgaatctcgg gtacgtgttc cggacatggg ctcttctccg 1560 gtageggegg agetteeaca teegageest ggteecatge etecagegge teatggtege 1620 teggeagete ettgeteeta acagtggagg ecagaettag geacageaca atgeceacca 1680 ccaccagtgt gccgcacaag gccgtggcgg tagggtatgt gtctgaaaat gagctcggag 1740 attgggctcg caccgctgac gcagatggaa gacttaaggc agcggcagaa gaagatgcag 1800 gcagctgagt tgttgtattc tgataagagt cagaggtaac tcccgttgcg gtgctgttaa 1860

cggtggaggg cagtgtagtc tgagcagtac tcgttgctgc cgcgcgccc accagacata 1920 atagctgaca gactaacaga ctgttccttt ccatgggtct tttctgcagt caccgtcgtc 1980 gacctaagaa ttcagactcg agcaagtcta gaaaggcgcg ccaagatatc aaggatccac 2040 tacgcgttag agctcgctga tcagcctcga ctgtgccttc tagttgccag ccatctgttg 2100 tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact gtcctttcct 2160 aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt ctggggggtg 2220 gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat gctggggagc 2280 tetteegett cetegeteae tgactegetg egeteggteg tteggetgeg gegageggta 2340 tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa cgcaggaaag 2400 aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 2460 tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg 2520 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 2580 egeteteetg tteegaceet geegettace ggatacetgt cegeetttet ceetteggga 2640 agegtggege ttteteaatg eteaegetgt aggtatetea gtteggtgta ggtegttege 2700 tecaagetgg getgtgtgca egaaceeeee gtteageeeg acegetgege ettateeggt 2760 aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact 2820 ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg 2880 cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt 2940 accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt 3000 ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 3060 ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg 3120 gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt 3180 aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt 3240 gaggcaccta teteagegat etgtetattt egtteateca tagttgeetg acteeeegte 3300 gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg 3360 cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc cggaagggcc 3420 gagegeagaa gtggteetge aactttatee geeteeatee agtetattaa ttgttgeegg 3480 gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca 3540 ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg ttcccaacga 3600 tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc cttcggtcct 3660

ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg 3720
cataattctc ttactgtcat gccatccgta agatgcttt ctgtgactgg tgagtactca 3780
accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata 3840
cggggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct 3900
tcgggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaacccact 3960
cgtgcaccca actgatcttc agcatcttt actttcacca gcgtttctgg gtgagcaaaa 4020
acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc 4080
atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga 4140
tacatatttg aatgtatta gaaaaataaa caaatagggg ttccgcgcac atttcccga 4200
aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg 4260
cgtatcacga ggccctttcg tc 4282

<210> 6

<211> 6299 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: pNS34a <220> <221> CDS <222> (1990)..(4047) <400> 6 egegegttte ggtgatgaeg gtgaaaacet etgacacatg eagetceegg agaeggteae 60 agettqtctq taaqeqqatq ceqqqaqcaq acaaqeeeqt caqqqqqqtqt 120 tggegggtgt cggggctggc ttaactatgc ggcatcagag cagattgtac tgagagtgca 180 ccatatgaag ctttttgcaa aagcctaggc ctccaaaaaa gcctcctcac tacttctgga 240 atageteaga ggeegaggeg geeteggeet etgeataaat aaaaaaaatt agteageeat 300 ggggcggaga atgggcggaa ctgggcgggg agggaattat tggctattgg ccattgcata 360 cgttgtatct atatcataat atgtacattt atattggctc atgtccaata tgaccgccat 420 gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 480 gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 540 ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 600 ggactttcca ttgacgtcaa tgggtggagt atttacqgta aactgcccac ttggcaqtac 660 atcaagtgta tcatatgcca agtccgccc ctattgacgt caatgacggt aaatggcccq 720

cctggcatta tgcccagtac atgacettac gggactttcc tacttggcag tacatetacg 780 tattagtcat cgctattacc atggtgatgc ggttttggca gtacaccaat gggcgtggat 840 agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 900 tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc ccgttgacgc 960 aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctcgt ttagtgaacc 1020 gtcagatege etggagaege catecaeget gttttgaeet ecatagaaga caeegggaee 1080 gatecageet cegeggeegg gaacggtgea ttggaacgeg gatteceegt gecaaqagtg 1140 acgtaagtac cgcctataga ctctataggc acaccccttt ggctcttatg catgctatac 1200 tgtttttggc ttggggccta tacacccccg ctccttatgc tataggtgat ggtatagctt 1260 agcctatagg tgtgggttat tgaccattat tgaccactcc cctattggtg acgatacttt 1320 ccattactaa tccataacat ggctctttgc cacaactatc tctattggct atatgccaat 1380 actetyteet teagagacty acaeggacte tytattttta caggatygyg tecattatt 1440 atttacaaat tcacatatac aacaacgccg tcccccgtgc ccgcagtttt tattaaacat 1500 agcgtgggat ctccgacatc tcgggtacgt gttccggaca tgggctcttc tccggtagcg 1560 geggagette cacateegag ceetggteee ateegteeag eggeteatgg tegeteggea 1620 gctccttgct cctaacagtg gaggccagac ttaggcacag cacaatgccc accaccacca 1680 gtgtgccgca caaggccgtg gcggtagggt atgtgtctga aaatgagctc ggagattggg 1740 ctcgcacctg gacgcagatg gaagacttaa ggcagcggca gaagaagatg caggcagctg 1800 agttgttgta ttctgataag agtcagaggt aactcccgtt gcggtgctgt taacggtgga 1860 gggcagtgta gtctgagcag tactcgttgc tgccgcgcgc gccaccagac ataatagctg 1920 acagactaac agactgttcc tttccatggg tcttttctgc agtcaccgtc qtcqacctaa 1980 gaattcacc atg gcg ccc atc acg gcg tac gcc cag cag aca agg ggc ctc 2031 Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu cta ggg tgc ata atc acc agc cta act ggc cgg gac aaa aac caa gtg 2079 Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 20 gag ggt gag gtc cag att gtg tca act gct gcc caa acc ttc ctq qca 2127 Glu Gly Glu Val Gln Ile Val Ser Thr Ala Ala Gln Thr Phe Leu Ala acg tgc atc aat ggg gtg tgc tgg act gtc tac cac ggg gcc gga acg 2175 Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr agg acc atc gcg tca ccc aag ggt cct gtc atc cag atg tat acc aat 2223

Arg	Thr	Ile 65	Ala	Ser	Pro	Lys	Gly 70	Pro	Val	Ile	Gln	Met 75	Tyr	Thr	Asn	
				ctt Leu												2271
_			_	act Thr	_			_	_			_	-	_		2319
				att Ile 115												2367
		_		cgg Arg					_				_			2415
		-	_	ccc Pro	_			_						_		2463
				gga Gly												2511
				acc Thr												2559
				ccc Pro 195												2607
				aaa Lys												2655
				cta Leu												2703
				atg Met												2751
-			_	aca Thr					_			_				2799
tac Tyr	ggc Gly	aag Lys	ttc Phe	ctt Leu 275	gcc Ala	gac Asp	ggc Gly	ggg Gly	tgc Cys 280	tcg Ser	ggg ggg	ggc Gly	gct Ala	tat Tyr 285	gac Asp	2847
ata Ile	ata Ile	att Ile	tgt Cys 290	gac Asp	gag Glu	tgc Cys	cac His	tcc Ser 295	acg Thr	gat Asp	gcc Ala	aca Thr	tcc Ser 300	atc Ile	ttg Leu	2895

				_		_		_	~ ~				gcg Ala	_	_	2943
_			_		_			_			_		gtg Val			2991
													atc Ile			3039
		_	_					_		_			aga Arg			3087
		_			_	_	_	_	_	_		_	gca Ala 380	_	-	3135
													ctt Leu			3183
													acc Thr			3231
	_						_		-	_			gac Asp	_		3279
_	_	_		_		_	-		_		_		acc Thr			3327
													act Thr 460			3375
							_					_	ttt Phe		_	3423
_		-	_				_		_	_		_	ctc Leu	_		3471
_		_	_		_	_					_		gcc Ala			3519
													ccc Pro			3567
													ctc Leu 540			3615

ata gat gcc cac ttt cta tcc cag aca aag cag agt ggg gag aac ctt 3663 Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu 545 550 555
cct tac ctg gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc 3711 Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala 560 570
CCT CCC CCa tcg tgg gac Cag atg tgg aag tgt ttg att cgc ctc aag 3759 Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys 575 580 585 590
CCC acc ctc cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt 3807 Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val 595 600 605
cag aat gaa atc acc ctg acg cac cca gtc acc aaa tac atc atg aca 3855 Gln Asn Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr 610 615 620
tgc atg tcg gcc gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt 3903 Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val 625 630 635
ggc ggc gtc ctg gct gct ttg gcc gcg tat tgc ctg tca aca ggc tgc 3951 Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys 640 645 650
gtg gtc ata gtg ggc agg gtc gtc ttg tcc ggg aag ccg gca atc ata 3999 Val Val Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile 655 660 665 670
cct gac agg gaa gtc ctc tac cga gag ttc gat gag atg gaa gag tgc 4047 Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys 675 680 685
taggatccac tacgcgttag agctcgctga tcagcctcga ctgtgccttc tagttgccag 4107
ccatctgttg tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact 4167
gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 4227
ctggggggtg gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat 4287
gctggggagc tettccgctt cctcgctcac tgactcgctg egctcggtcg ttcggctgcg 4347
gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa 4407
cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 4467
gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc 4527
aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 4587
ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct 4647
cccttcggga agcgtggcgc tttctcaatg ctcacgctgt aggtatctca gttcggtgta 4707
ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc 4767

cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc 4827 agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 4887 gaagtggtgg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 4947 gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 5007 tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 5067 agaagateet ttgatetttt etaeggggte tgaegeteag tggaacgaaa aeteaegtta 5127 agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 5187 atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg 5247 cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 5307 acteccegte gtgtagataa ctacgatacg ggagggetta ccatetggce ccagtgetge 5367 aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc 5427 cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 5487 ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc 5547 cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg 5607 ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaaag cggttagctc 5667 cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 5727 ggcagcactg cataattoto ttactgtcat gccatccgta agatgctttt ctgtgactgg 5787 tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 5847 ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaaagtgc tcatcattgg 5907 aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 5967 gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg 6027 gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg 6087 ttgaatactc atactettee tttttcaata ttattgaage atttateagg gttattgtet 6147 catgagegga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgegcac 6207 atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta 6267 taaaaatagg cgtatcacga ggccctttcg tc 6299

<210> 7

<211> 686

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: pNS34a

<400> 7

Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20 25 30

Glu Val Gln Ile Val Ser Thr Ala Ala Gln Thr Phe Leu Ala Thr Cys 35 40 45

Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr 50 60

Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp 65 70 75 80

Gln Asp Leu Val Gly Trp Pro Ala Ser Gln Gly Thr Arg Ser Leu Thr 85 90 95

Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala 100 105 110

Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu 115 120 125

Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu 130 135 140

Leu Cys Pro Ala Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys 145 150 155 160

Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu 165 170 175

Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro 180 185 190

Val Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly
195 200 205

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr 210 215 220

Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly 225 230 235 240

Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
245 250 255

Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly
260 265 270

Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile 275 280 285

Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile 290 295 300

Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn 330 Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys 425 Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp 520 His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr 585 Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met 615 620

Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly 625 630 635 640

Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val 645 650 655

Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp
660 665 670

Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys
675 680 685

<210> 8

<211> 19912

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: pd.deltaNS3NS5

<220>

<221> CDS

<222> (12745)..(18057)

<400> 8

atogatocta coccttgogo taaagaagta tatgtgocta ctaacgottg totttgtoto 60 tgtcactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120 eggateaacg ttettaatat egetgaatet tecacaattg atgaaagtag etaggaagag 180 gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240 tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300 actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgtaatt gcaaactact 360 tacaatatca gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420 ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgcttttt gcgtcttgta 480 tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540 atageteatg aatgtggete tettgattge tgtteegtta tgtgtaatea tecaacataa 600 ataggttagt tcagcagcac ataatgctat tttctcacct gaaggtcttt caaacctttc 660 cacaaactga cgaacaagca ccttaggtgg tgttttacat aatatatcaa attgtggcat 720 gettagegee gatettgtgt geaattgata tetagtttea aetactetat ttatettgta 780 tettgeagta tteaaacaeg etaactegaa aaactaaett taattgteet gtttgteteg 840 cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attgqtactq 900 cggtatcttc atttcatatt ttaaaaatgc acctttgctg cttttcctta atttttagac 960 ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtgatc 1020

tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080 agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140 cttatagcgt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200 ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260 ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320 gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380 ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agetegtaag egtegttace caattgetta etttagttee accagategt caagetgett 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740 tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaaatg 1860 caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340 ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctg aagacgcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tecacaaaca ttgeteaaaa gtatetettt getatatate tetgtgetat atecetatat 2520 aacctaccca tecacettte geteettgaa ettgeateta aactegaeet etacateaac 2580 aggetteeaa tgetetteaa attttaetgt caagtagace cataeggetg taatatgetg 2640 ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820

gaaacatgct gcttaaaact ccaagcggta ggagaccgat aaaggttaat aggacagccg 2880 tattatetee geeteagttt gatetteege tteagaetge eattttteae ataatgaate 2940 tatttcaccc cacaatcett catcegeete egeatettgt teegttaaac tattgaette 3000 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggcggtagct cctgatctcc 3060 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3120 gcacctaata acattettea aggeggagaa gtttgggeca gatgeecaat atgettgaca 3180 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3240 tcgaagataa gagaagaatg cagtgacctt tgtattgaca aatggagatt ccatgtatct 3300 aaaaaatacg cctttaggcc ttctgatacc ctttcccctg cggtttagcg tgccttttac 3360 attaatatct aaaccctctc cgatggtggc ctttaactga ctaataaatg caaccgatat 3420 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3540 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3600 cccccttgtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3660 cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttattttc 3720 cattacagct gatgtcattg tatatcagcg ctgtaaaaat ctatctgtta cagaaggttt 3780 tegeggtttt tataaacaaa aetttegtta egaaategag caateacece agetgegtat 3840 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3900 ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaaac ctcaatctgt 3960 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 4020 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4080 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4140 tctactgtac gatacacttc cgctcaggtc cttgtccttt aacgaggcct taccactctt 4200 ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4260 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4320 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt tttttttt 4380 ttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatctttta 4440 agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcggtgg tactgttgga 4500 accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4560 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4620

agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4680 gtacaaacca aatgcggtgt tcttgtctgg caaagaggcc aaggacgcag atggcaacaa 4740 acceaaggag cetgggataa eggaggette ateggagatg atateaceaa acatgttget 4800 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4860 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4920 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4980 tggtggctca tgttgtaggg ccatgaaagc ggccattett gtgattettt gcacttetgg 5040 aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5100 aaagtaaata cctcccacta attctctaac aacaacgaag tcagtacctt tagcaaattg 5160 tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5220 ggcgtacaat tgaagttett tacggatttt tagtaaacet tgttcaggte taacactace 5280 ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5340 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5400 atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5460 aatggcttcg gctgtgattt cttgaccaac gtggtcacct ggcaaaacga cgatcttctt 5520 aggggcagac attacaatgg tatatcettg aaatatatat aaaaaaaaaa aaaaaaaaa 5580 aaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5640 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5700 acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5760 tatagtctag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5820 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5880 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5940 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 6000 gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6060 caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6120 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6180 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6240 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccqt 6300 taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6360 cactteecge gtttactgat tactagegaa getgegggtg catttttea agataaagge 6420

atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6480 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6540 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6600 tettactaca attttttgt etaaagagta atactagaga taaacataaa aaatgtagag 6660 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6720 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6780 aatattttag tagctcgtta cagtccggtg cgtttttggt tttttgaaag tgcgtcttca 6840 gagcgctttt ggttttcaaa agcgctctga agttcctata ctttctagag aataggaact 6900 teggaatagg aactteaaag egttteegaa aacgageget teegaaaatg caaegegage 6960 tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatatata 7020 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7080 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7140 gtatcgtatg cttccttcag cactaccctt tagctgttct atatgctgcc actcctcaat 7200 tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7260 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7320 cctggccacg gcagaagcac gcttatcgct ccaatttccc acaacattag tcaactccgt 7380 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7440 attititata gcaaagatig aataaggcgc attititctic aaagctitat tgtacgatct 7500 gactaagtta tcttttaata attggtattc ctgtttattg cttgaagaat tgccggtcct 7560 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7620 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7680 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7740 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7800 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7860 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7920 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7980 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 8040 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8100 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8160 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8220

gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8280 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8340 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8400 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8460 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8520 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8580 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 8640 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8700 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8760 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8820 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8880 gatettettg agateetttt tttetgegeg taatetgetg ettgeaaaca aaaaaaceac 8940 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactettttt ccgaaggtaa 9000 ctggcttcag cagagegeag ataccaaata ctgtccttct agtgtagecg tagttaggec 9060 accactteaa gaactetgta geacegeeta catacetege tetgetaate etgttaceag 9120 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9180 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9240 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9300 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9360 cgagggagct tccaggggga aacgectggt atetttatag teetgteggg tttegecace 9420 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9480 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9540 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9600 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9660 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9720 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9780 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9840 ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 9900 gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9960 agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gctcgttgag 10020

tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10080 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10140 accgatgaaa cgagagagga tgctcacgat acgggttact gatgatgaac atgcceggtt 10200 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10260 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10320 gcagcatect gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10380 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10440 tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10500 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10560 tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10620 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10680 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10740 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10800 gegectacaa tecatgecaa ecegtteeat gtgetegeeg aggeggeata aategeegtg 10860 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10920 tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10980 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 11040 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11100 ccgaaacgtt tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11160 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11220 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11280 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11340 ctcaagggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggtc 11400 ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaatct 11460 atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11520 attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcacaa 11580 caataccagt coggatcaac tggcaccatc totocogtag totoatotaa tttttcttcc 11640 ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11700 atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatttcgg gtttgttaac 11760 tegttecagt caggaatgtt ccaegtgaag etatetteca gcaaagtete caettettea 11820

tcaaattgtg gagaatactc ccaatgctct tatctatggg acttccggga aacacagtac 11	380
cgatacttcc caattcgtct tcagagetca ttgtttgttt gaagagacta atcaaagaat 11	940
cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgtagg 12	000
ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12	060
tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaatcc 12	120
tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggctt 12	180
tcgctcataa aaatgttatg acgttttgcc cgcaggcggg aaaccatcca cttcacgaga 12	240
ctgatctcct ctgccggaac accgggcatc tccaacttat aagttggaga aataagagaa 12	300
tttcagattg agagaatgaa aaaaaaaac ccttagttca taggtccatt ctcttagcgc 12	360
aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgatg 12	420
caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgta tctatctcat 12	480
tttcttacac cttctattac cttctgctct ctctgatttg gaaaaagctg aaaaaaagg 12	540
ttgaaaccag ttccctgaaa ttattcccct acttgactaa taagtatata aagacggtag 12	500
gtattgattg taattotgta aatotattto ttaaaottot taaattotao ttttatagtt 12	660
agtetttttt ttagttttaa aacaccaaga aettagttte gaataaacae acataaacaa 12	720
acaagcttac aaaacaaatt cacc atg gct gca tat gca gct cag ggc tat 12 Met Ala Ala Tyr Ala Ala Gln Gly Tyr 1 5	771
aag gtg cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt 12 Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly 10 15 20 25	819
gct tac atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg 12 Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly 30 35 40	867
gtg aga aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc 12 Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly 45 50 55	915
aag ttc ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata 12 Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile 60 65 70	963
att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att 13	011
Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile 75 80 85	
Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile 75 80 85	059

Leu Ala Thr A	Ala Thr Pro 110	Pro Gly Ser	Val Thr Val	Pro His Pro	Asn
atc gag gag g Ile Glu Glu V 1			Gly Glu Ile		
aag gct atc c Lys Ala Ile F 140	_	-			
tgt cat tca a Cys His Ser I 155					_
ttg ggc atc a Leu Gly Ile A 170		_			-
atc ccg acc a					
acc ggc tat a Thr Gly Tyr T			Val Ile Asp		
gtc acc cag a Val Thr Gln 7 220		_	_		
aca atc acg of Thr Ile Thr I 235			-		
agg act ggc a Arg Thr Gly A 250			_		
gag cgc ccc t Glu Arg Pro S		•	_	J J J	
gac gca ggc t Asp Ala Gly (Thr Pro Ala		
agg cta cga q Arg Leu Arg <i>l</i> 300		_			_
cat ctt gaa t His Leu Glu I 315					•
gcc cac ttt o Ala His Phe I 330					

ctg Leu		gcg Ala														13827
	_	tgg Trp	_	_	_		_	_	_		_		_			13875
ctc Leu		ggg Gly 380														13923
_		acc Thr	_	_			_					_		_	_	13971
_	_	gac Asp	_		_	_	_	_					_			14019
_	_	gct Ala	_	_	_			_	_				_		_	14067
		ggc Gly														14115
	-	gtc Val 460			_			_		_	_		-		_	14163
		ccg Pro														14211
		gcc Ala														14259
		cct Pro	Ala				Asn		Gln	Lys						14307
gcg Ala	aag Lys	cat His	atg Met 525	tgg Trp	aac Asn	ttc Phe	atc Ile	agt Ser 530	ggg Gly	ata Ile	caa Gln	tac Tyr	ttg Leu 535	gcg Ala	ggc Gly	14355
		acg Thr 540														14403
	_	gct Ala	_		_					_						14451
		ttg Leu														14499

				gtg Val 590												14547
				aag Lys												14595
				gct Ala												14643
Pro		_	_	gac Asp	_	_			_		_			_		14691
	_		_	gtc Val				_	_	_		_	_			14739
				gag Glu 670												14787
	_			gly ggg			_			_				_		14835
_	_	_	_	gcc Ala	_	_		_			_	_			_	14883
	_		_	agg Arg	_	_		_			_	_		_		14931
		_		ggt Gly					_			_			_	14979
				gac Asp 750		Lys	Thr	Trp	Leu	Lys	Ala	Lys	Leu			15027
				atc Ile												15075
				gac Asp												15123
				cat His												15171
				aac Asn												15219

											aac Asn					15267
											ata Ile					15315
_					_		_			-	aat Asn			_	_	15363
											ttg Leu 885					15411
											ctg Leu					15459
											gjà aaa					15507
											tcc Ser					15555
					_					_	agg Arg	_				15603
					_	_		_	_	_	cag Gln 965			_		15651
		_	_		_		_			_	tcc Ser		_	_		15699
			Āla	Asn	Leu	Leu	Trp		Gln	Glu	atg Met		Gly		Ile	15747
		Val					Lys				ctg Leu	Asp			gat Asp	15795
	Leu					Asp					tcc Ser					15843
Ile					Arg					Ala	ctg Leu 1045				gcg Ala	15891
	Pro			Asn		Pro			Glu		tgg Trp			Pro	gac Asp 1065	15939

tac gaa cca cct gtg gtc Tyr Glu Pro Pro Val Val 1070			
cct cct gtg cct ccg cct Pro Pro Val Pro Pro Pro 1085			
tca acc cta tct act gcc Ser Thr Leu Ser Thr Ala 1100			
agc tcc tca act tcc ggc Ser Ser Ser Thr Ser Gly 1115			
gag ccc gcc cct tct ggc Glu Pro Ala Pro Ser Gly 1130 1135	Cys Pro Pro Asp	Ser Asp Ala Glu Ser T	
tcc tcc atg ccc ccc ctg Ser Ser Met Pro Pro Leu 1150			
gac ggg tca tgg tca acg Asp Gly Ser Trp Ser Thr 1165	gtc agt agt gag Val Ser Ser Glu 1170	gcc aac gcg gag gat g Ala Asn Ala Glu Asp V 1175	gtc 16275 <i>V</i> al
gtg tgc tgc tca atg tct Val Cys Cys Ser Met Ser 1180			
tgc gcc gcg gaa gaa cag Cys Ala Ala Glu Glu Gln 1195			
ttg cta cgt cac cac aat Leu Leu Arg His His Asn 1210 1215	Leu Val Tyr Ser	Thr Thr Ser Arg Ser A	gct 16419 Ala 225
tgc caa agg cag aag aaa Cys:Gln Arg Gln Lys Lys 1230		Arg Leu Gln Val Leu A	
agc cat tac cag gac gta Ser His Tyr Gln Asp Val 1245			
gtg aag gct aac ttg cta Val Lys Ala Asn Leu Leu 1260			
cca cac tca gcc aaa tcc Pro His Ser Ala Lys Ser 1275			
tgc cat gcc aga aag gcc Cys His Ala Arg Lys Ala 1290 1295	Val Thr His Ile	Asn Ser Val Trp Lys A	gac 16659 Asp 305

ctt ctg gaa gac aat g Leu Leu Glu Asp Asn v 1310	Val Thr Pro Ile		
aac gag gtt ttc tgc q Asn Glu Val Phe Cys v 1325		Lys Gly Gly Arg I	
cgt ctc atc gtg ttc o Arg Leu Ile Val Phe 1 1340	ccc gat ctg ggc Pro Asp Leu Gly 1345	gtg cgc gtg tgc g Val Arg Val Cys (1350	gaa aag atg 16803 Glu Lys Met
gct ttg tac gac gtg g Ala Leu Tyr Asp Val v 1355			
tcc tac gga ttc caa g Ser Tyr Gly Phe Gln 1 1370			
caa gcg tgg aag tcc a Gln Ala Trp Lys Ser 1 1390	Lys Lys Thr Pro		
cgc tgc ttt gac tcc a Arg Cys Phe Asp Ser 1 1405	aca gtc act gag Thr Val Thr Glu 1410	Ser Asp Ile Arg	ncg gag gag 16995 Thr Glu Glu H15
gca atc tac caa tgt o Ala Ile Tyr Gln Cys o 1420			
aag too oto acc gag a Lys Ser Leu Thr Glu a 1435	agg ctt tat gtt Arg Leu Tyr Val 1440	ggg ggc cct ctt a Gly Gly Pro Leu 1 1445	cc aat tca 17091 Thr Asn Ser
agg ggg gag aac tgc g Arg Gly Glu Asn Cys (1450			
aca act agc tgt ggt a Thr Thr Ser Cys Gly a 1470	Asn Thr Leu Thr		la Arg Ala
gcc tgt cga gcc gca g Ala Cys Arg Ala Ala (1485	ggg ctc cag gac Gly Leu Gln Asp 1490	Cys Thr Met Leu \	gtg tgt ggc 17235 Val Cys Gly 195
gac gac tta gtc gtt a Asp Asp Leu Val Val : 1500	atc tgt gaa agc Ile Cys Glu Ser 1505	gcg ggg gtc cag g Ala Gly Val Gln (1510	gag gac gcg 17283 Slu Asp Ala
gcg agc ctg aga gcc f Ala Ser Leu Arg Ala 1 1515	ttc acg gag gct Phe Thr Glu Ala 1520	atg acc agg tac t Met Thr Arg Tyr S 1525	cc gcc ccc 17331 Ser Ala Pro
cct ggg gac ccc cca o Pro Gly Asp Pro Pro (1530	caa cca gaa tac Gln Pro Glu Tyr 535	gac ttg gag ctc a Asp Leu Glu Leu 1 1540	ta aca tca 17379 le Thr Ser 1545

tgc tcc tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val 1550 1560	17427
tac tac ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp 1565 1570 1575	17475
gag aca gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile 1580 1585 1590	17523
atg ttt gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe 1595 1600 1605	17571
ttt agc gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys 1610 1625	17619
gag atc tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro 1630 1635 1640	17667
atc att caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr 1645 1650 1655	17715
tct cca ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly 1660 1665 1670	17763
gta ccg ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala 1675 1680 1685	17811
agg ctt ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc Arg Leu Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu 1690 1695 1700 1705	17859
ttc aac tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala 1710 1715 1720	17907
gct ggc cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg Ala Gly Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly 1725 1730 1735	17955
gga gac att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg Gly Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp 1740 1745 1750	18003
ttt tgc cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ccc Phe Cys Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro 1755 1760 1765	18051
aac cga tgaaggttgg ggtaaacact ccggcctaaa aaaaaaaaa aatctagaac Asn Arg 1770	18107

 $\mathbb{C}_{n}^{2}\neq$

ccgagtcgac tttgttccca ctgtactttt agctcgtaca aaatacaata tacttttcat 18167 ttctccgtaa acaacatgtt ttcccatgta atatcctttt ctatttttcg ttccgttacc 18227 aactttacac atactttata tagctattca cttctataca ctaaaaaact aagacaattt 18287 taattttgct gcctgccata tttcaatttg ttataaattc ctataattta tcctattagt 18347 agctaaaaaa agatgaatgt gaatcgaatc ctaagagaat tggatctgat ccacaggacg 18407 ggtgtggtcg ccatgatcgc gtagtcgata gtggctccaa gtagcgaagc gagcaggact 18467 gggcggcggc caaagcggtc ggacagtgct ccgagaacgg gtgcgcatag aaattgcatc 18527 aacgcatata gcgctagcag cacgccatag tgactggcga tgctgtcgga atggacgata 18587 tecegeaaga ggeeeggeag taceggeata accaageeta tgeetacage atceagggtg 18647 acggtgccga ggatgacgat gagcgcattg ttagatttca tacacqqtqc ctqactqcqt 18707 tagcaattta actgtgataa actaccgcat taaagctttt tctttccaat ttttttttt 18767 tegteattat aaaaateatt aegacegaga tteeegggta ataactgata taattaaatt 18827 gaagetetaa titigtgagti tagtatacat geatitaeti ataatacagt tititagtit 18887 tgctggccgc atcttctcaa atatgcttcc cagcctgctt ttctgtaacg ttcaccctct 18947 accttagcat cccttccctt tgcaaatagt cctcttccaa caataataat gtcagatcct 19007 gtagagacca catcatccac ggttctatac tgttgaccca atgcgtctcc cttgtcatct 19067 aaacccacac cgggtgtcat aatcaaccaa tcgtaacctt catctcttcc acccatgtct 19127 ctttgagcaa taaagccgat aacaaaatct ttgtcgctct tcgcaatgtc aacagtaccc 19187 ttagtatatt ctccagtaga tagggagccc ttgcatgaca attctgctaa catcaaaagg 19247 cctctaggtt cctttgttac ttcttctgcc gcctgcttca aaccgctaac aatacctggg 19307 cccaccacac cgtgtgcatt cgtaatgtct gcccattctg ctattctgta tacacccgca 19367 gagtactgca atttgactgt attaccaatg tcagcaaatt ttctgtcttc gaagagtaaa 19427 aaattgtact tggcggataa tgcctttagc ggcttaactg tgccctccat ggaaaaatca 19487 gtcaagatat ccacatgtgt ttttagtaaa caaattttgg gacctaatgc ttcaactaac 19547 tocagtaatt cottggtggt acgaacatcc aatgaagcac acaagtttgt ttgcttttcg 19607 tgcatgatat taaatagctt ggcagcaaca ggactaggat gagtagcagc acgttcctta 19667 tatgtagett tegacatgat ttatettegt tteetgeagg tttttgttet gtgcagttqq 19727 gttaagaata ctgggcaatt tcatgtttct tcaacactac atatgcgtat atataccaat 19787 ctaagtctgt gctccttcct tcgttcttcc ttctgttcgg agattaccga atcaaaaaaa 19847 tttcaaggaa accgaaatca aaaaaaagaa taaaaaaaaa atgatgaatt gaaaagctta 19907

tcgat 19912

<210> 9

<211> 1771

<212> PRT

<213> Artificial Sequence

2205

<223> Description of Artificial Sequence: pd.deltaNS3NS5

<400> 9

i ...

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly 35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val 180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 215 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235 240

Ala Val Ser Arg Thr Gln Arg Gly Arg Thr Gly Arg Gly Lys Pro
245 250 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr 280 Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 295 Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 310 315 Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro 395 Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 450 455 Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr 505 Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 520 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 550 Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val 565 570

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala 580 Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu 600 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 615 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val 630 Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val 650 Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val 695 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 855 Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 890

Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
900 905 910

- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Île Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230

- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro 1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550

- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 \$1740
- Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala 745 1750 1755 1760
- Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
- <210> 10
- <211> 19798
- <212> DNA
- <213> Artificial Sequence
- <220>
- <223> Description of Artificial Sequence:
 pd.deltaNS3NS5.pj
- <220>
- <221> CDS
- <222> (12679)..(17991)

<400> 10 atogatocta cocottgogo taaagaagta tatgtgoota otaacgottg totttgtoto 60 tgtcactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120 cggatcaacg ttcttaatat cgctgaatct tccacaattg atgaaagtag ctaggaagag 180 gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240 tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcattttttt 300 actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgtaatt gcaaactact 360 tacaatatca gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420 ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgcttttt gcgtcttgta 480 tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540 atageteatg aatgtggete tettgattge tgtteegtta tgtgtaatea teeaacataa 600 ataggttagt tcagcagcac ataatgctat tttctcacct gaaggtcttt caaacctttc 660 cacaaactga cgaacaagca ccttaggtgg tgttttacat aatatatcaa attgtggcat 720 gettagegee gatettgtgt geaattgata tetagtttea actaetetat ttatettgta 780 tettgeagta tteaaacaeg etaactegaa aaactaactt taattgteet gtttgteteg 840 cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attggtactg 900 cggtatcttc atttcatatt ttaaaaatgc acctttgctg cttttcctta atttttagac 960 ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtgatc 1020 tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080 agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140 cttatagcgt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200 ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260 ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320 gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380 ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agctogtaag ogtogttaco caattgotta otttagttoo accagatogt caagotgott 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740

tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaaatg 1860 caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340 ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctg aagacgcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tccacaaaca ttgctcaaaa gtatctcttt gctatatatc tctgtgctat atccctatat 2520 aacctaccca tecacettte geteettgaa ettgeateta aactegacet etacateaac 2580 aggetteeaa tgetetteaa attttaetgt caagtagace cataeggetg taatatgetg 2640 ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820 tattatctcc gcctcagttt gatcttccgc ttcagactgc catttttcac ataatgaatc 2880 tatttcaccc cacaatcett catcegeete egeatettgt teegttaaac tattgaette 2940 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggcggtagct cctgatctcc 3000 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060 gcacctaata acattettea aggeggagaa gtttgggeea gatgeecaat atgettgaea 3120 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180 tcgaagataa gagaagaatg cagtgacctt tgtattgaca aatggagatt ccatgtatct 3240 aaaaaatacg cctttaggcc ttctgatacc ctttcccctg cggtttagcg tgccttttac 3300 attaatatet aaaccetete egatggtgge etttaaetga etaataaatg caaccgatat 3360 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3480 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540

cccccttgtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3600 cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttattttc 3660 cattacagct gatgtcattg tatatcagcg ctgtaaaaat ctatctgtta cagaaggttt 3720 tegeggtttt tataaacaaa actttegtta egaaategag caateacece agetgegtat 3780 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840 ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 3960 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4020 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4080 totactgtac gatacacttc cgctcaggtc cttgtccttt aacgaggcct taccactctt 4140 ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt ttttttttt 4320 tttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatcttttta 4380 agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcggtgg tactgttgga 4440 accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4500 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560 agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4620 gtacaaacca aatgcggtgt tcttgtctgg caaagaggcc aaggacgcag atggcaacaa 4680 acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgct 4740 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4800 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4860 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920 tggtggctca tgttgtaggg ccatgaaagc ggccattctt gtgattcttt gcacttctgg 4980 aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5040 aaagtaaata cctcccacta attctctaac aacaacgaag tcagtacctt tagcaaattg 5100 tggcttgatt ggagataagt ctaaaagag gtcggatgca aagttacatg gtcttaagtt 5160 ggcgtacaat tgaagttett tacggatttt tagtaaacet tgttcaggte taacactace 5220 ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5340

atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400 aatggetteg getgtgattt ettgaceaac gtggteaeet ggeaaaaega egatettett 5460 aaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640 acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700 tatagtctag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 5940 gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000 caaaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttt 6120 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240 taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6300 cacttcccgc gtttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6360 atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6540 tcttactaca atttttttgt ctaaagagta atactagaga taaacataaa aaatgtagag 6600 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6660 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720 aatattttag tagetegtta eagteeggtg egtttttggt tttttgaaag tgegtettea 6780 gagegetttt ggtttteaaa agegetetga agtteetata etttetagag aataggaact 6840 teggaatagg aactteaaag egttteegaa aaegageget teegaaaatg caaegegage 6900 tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatatata 6960 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7080 gtategtatg ctteetteag cactaceett tagetgttet atatgetgee acteeteaat 7140

tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7200 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260 cctggccacg gcagaagcac gcttatcgct ccaatttccc acaacattag tcaactccgt 7320 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380 attitttata gcaaagattg aataaggcgc attittcttc aaagctitat tgtacgatct 7440 gactaagtta tottttaata attggtatto ctgtttattg cttgaagaat tgccggtcct 7500 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7560 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7680 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7740 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7860 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 7980 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8040 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8160 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8220 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8280 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8400 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 8580 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8700 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8760 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8880 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 8940

ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 9000 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 9060 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9240 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300 cgagggaget tecaggggga aacgeetggt atetttatag teetgteggg tttegeeace 9360 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9420 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9480 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaaqagc 9600 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9660 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9720 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780 ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 9840 gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9900 agegtggteg tgaagegatt cacagatgte tgeetgttea teegegteea getegttgag 9960 tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080 accgatgaaa cgagagga tgctcacgat acgggttact gatgatgaac atgcccggtt 10140 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10260 gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10320 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380 tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500 tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10620 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10680 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10740

gcgcctacaa tccatgccaa cccgttccat gtgctcgccg aggcggcata aatcgccgtg 10800 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860 tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 10980 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11040 ccgaaacgtt tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11160 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11280 ctcaagggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggtc 11340 ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaatct 11400 atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460 attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcacaa 11520 caataccagt ccggatcaac tggcaccatc tctcccgtag tctcatctaa tttttcttcc 11580 ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640 atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatttcgg gtttgttaac 11700 tegtteeagt caggaatgtt ceaegtgaag etatetteea geaaagtete eaettettea 11760 tcaaattgtg gagaatactc ccaatgctct tatctatggg acttccggga aacacagtac 11820 cgatacttcc caattcgtct tcagagctca ttgtttgttt gaagagacta atcaaagaat 11880 cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgtagg 11940 ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000 tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaatcc 12060 tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggctt 12120 tegeteataa aaatgttatg aegttttgee egcaggeggg aaaccateea etteaegaga 12180 ctgatctcct ctgccggaac accgggcatc tccaacttat aagttggaga aataagagaa 12240 tttcagattg agagaatgaa aaaaaaaaac ccttagttca taggtccatt ctcttagcgc 12300 aactacagag aacagggca caaacaggca aaaaacgggc acaacctcaa tggagtgatg 12360 caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgta tctatctcat 12420 tttcttacac cttctattac cttctgctct ctctgatttg gaaaaagctg aaaaaaaagg 12480 ttgaaaccag ttccctgaaa ttattcccct acttgactaa taagtatata aagacggtag 12540

gtattgattg taatto	etgta aatetattt	c ttaaacttct	taaattctac ttttatagtt	12600
agtcttttt ttagtt	ttaa aacaccaag	ga acttagtttc	gaataaacac acataaacaa	12660
acaagcttac aaaaca			cag ggc tat aag gtg Gln Gly Tyr Lys Val 10	12711
			ggc ttt ggt gct tac Gly Phe Gly Ala Tyr 25	12759
		Pro Asn Ile	agg acc ggg gtg aga Arg Thr Gly Val Arg 40	12807
			acc tac ggc aag ttc Thr Tyr Gly Lys Phe 55	12855
			gac ata ata att tgt Asp Ile Ile Ile Cys 75	12903
			ttg ggc att ggc act Leu Gly Ile Gly Thr 90	12951
			ctg gtt gtg ctc gcc Leu Val Val Leu Ala 105	12999
		Thr Val Pro	cat ccc aac atc gag His Pro Asn Ile Glu 120	13047
			ttt tac ggc aag gct Phe Tyr Gly Lys Ala 135	13095
		Gly Arg His	ctc atc ttc tgt cat Leu Ile Phe Cys His 155	13143
Ser Lys Lys C			ctg gtc gca ttg ggc Leu Val Ala Leu Gly 170	13191
			gtg tcc gtc atc ccg Val Ser Val Ile Pro 185	13239
		Ala Thr Asp	gcc ctc atg acc ggc Ala Leu Met Thr Gly 200	13287
tat acc ggc gac t Tyr Thr Gly Asp F 205	ttc gac tcg gtg Phe Asp Ser Val 210	g ata gac tgc l Ile Asp Cys	aat acg tgt gtc acc Asn Thr Cys Val Thr 215	13335

cag ac Gln Th 220	_	_		_		-									13383
acg ct Thr Le															13431
ggc ag Gly Ar		_					_			_	_			-	13479
ccc to Pro Se		_		_	_		_		_		_		_	_	13527
ggc to Gly Cy 28	rs Ala														13575
cga go Arg Al 300															13623
gaa tt Glu Ph															13671
ttt ct Phe Le															13719
gcg ta Ala Ty															13767
tgg ga Trp As	p Gln	_		_	_	_		_		_					13815
ggg co Gly Pr 380															13863
acc ct Thr Le															13911
gac ct Asp Le															13959
gct gc Ala Al															14007
ggc ag Gly Ar 44	g Val	gtc Val	ttg Leu	tcc Ser	999 Gly 450	aag Lys	ccg Pro	gca Ala	atc Ile	ata Ile 455	cct Pro	gac Asp	agg Arg	gaa Glu	14055

_			_		ttc Phe 465	_		_	-		-		_			14103
_					Gly 999	_	_		_		_		-	_	_	14151
_				_	cag Gln				_	_	_		_		_	14199
	_	-	_		aac Asn										-	14247
	_				atc Ile	_					-			_		14295
_	_				ccc Pro 545	_		_		_	_	_			_	14343
-	_		_		cta Leu			_								14391
					gct Ala											14439
					ggc Gly											14487
					ata Ile											14535
			Leu	Val	gca Ala 625	Phe		Ile	Met	Ser	Gly					14583
					aat Asn										gcc Ala	14631
					gtc Val											14679
					gtg Val											14727
					gtt Val											14775

	cgc gtc act Arg Val Thr 705	_		_	•	14823
	cga ctg cac Arg Leu His 720					14871
	tcc tgg cta Ser Trp Leu 735		Trp Asp T			14919
	ttt aag acc Phe Lys Thr		-	_		14967
	ccc ttt gtg Pro Phe Val		Arg Gly T			15015
	ggc atc atg Gly Ile Met 785	_	•			15063
	gtc aaa aac Val Lys Asn 800					15111
	atg tgg agt Met Trp Ser 815		Pro Ile A	_	_	15159
	acc ccc ctt Thr Pro Leu					15207
	gca gag gaa Ala Glu Glu		Ile Arg G		_	15255
His Tyr Val	acg ggt atg Thr Gly Met 865	Thr Thr Asp	Asn Leu L	ys Cys Pro		15303
	ccc gaa ttt Pro Glu Phe 880					15351
agg ttt gcg Arg Phe Ala	ccc ccc tgc Pro Pro Cys 895	aag ccc ttg Lys Pro Leu 900	Leu Arg G	ag gag gta lu Glu Val 905	tca ttc Ser Phe	15399
	ctc cac gaa Leu His Glu					15447
ccc gaa ccg Pro Glu Pro 925	gac gtg gcc Asp Val Ala	gtg ttg acg Val Leu Thr 930	Ser Met L	tc act gat eu Thr Asp 35	ccc tcc Pro Ser	15495

cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940 945 950 955	15543
ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	15735
gtg gcg gag gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctg Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020 1025 1030 1035	15783
cgg aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cgg ccg Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040 1045 1050	15831
gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215 ;

tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935

tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly 1440 1445 1450	17031
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1455 1460 1465	17079
age tgt ggt aac acc etc act tge tac atc aag gee egg gea gee tgt Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys 1470 1475 1480	17127
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp 1485 1490 1495	17175
tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser 1500 1505 1510 1515	17223
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly 1520 1525 1530	17271
gac ccc cca caa cca gaa tac gac ttg gag ctc ata aca tca tgc tcc Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser 1535 1540 1545	17319
tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr 1550 1560	17367
ctc acc cgt gac cct aca acc ccc ctc gcg aga gct gcg tgg gag aca Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr 1565 1570 1575	17415
gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe 1580 1585 1590 1595	17463
gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser 1600 1605 1610	17511
gtc ctt ata gcc agg gac cag ctt gaa cag gcc ctc gat tgc gag atc Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile 1615 1620 1625	17559
tac ggg gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile 1630 1635 1640	17607
caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro 1645 1650 1655	17655

ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg 17703 Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro 1660 1665 1670 1675	
ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt 17751 Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu 1680 1685 1690	
ctg gcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac 17799 Leu Ala Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn 1695 1700 1705	
tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc 17847 Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly 1710 1715 1720	
cag ctg gac ttg tcc ggc tgg ttc acg gct ggc tac agc ggg gga gac 17895 Gln Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp 1725 1730 1735	
att tat cac age gtg tct cat gcc egg ece ege tgg ate tgg ttt tgc 17943 Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys 1740 1745 1750 1755	
cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga 17991 Leu Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1760 1765 1770	
tgaatagtcg actttgttcc cactgtactt ttagctcgta caaaatacaa tatacttttc 18051	
atttctccgt aaacaacatg ttttcccatg taatatcctt ttctattttt cgttccgtta 18111	
ccaactttac acatacttta tatagctatt cacttctata cactaaaaaa ctaagacaat 18171	
tttaattttg ctgcctgcca tatttcaatt tgttataaat tcctataatt tatcctatta 18231	
gtagctaaaa aaagatgaat gtgaatcgaa teetaagaga attggatetg atccacagga 18291	
cgggtgtggt cgccatgatc gcgtagtcga tagtggctcc aagtagcgaa gcgagcagga 18351	
ctgggcggcg gccaaagcgg tcggacagtg ctccgagaac gggtgcgcat agaaattgca 18411	
tcaacgcata tagcgctagc agcacgccat agtgactggc gatgctgtcg gaatggacga 18471	
tatcccgcaa gaggcccggc agtaccggca taaccaagcc tatgcctaca gcatccaggg 18531	
tgacggtgcc gaggatgacg atgagcgcat tgttagattt catacacggt gcctgactgc 18591	
gttagcaatt taactgtgat aaactaccgc attaaagctt tttctttcca atttttttt 18651	
tttcgtcatt ataaaaatca ttacgaccga gattcccggg taataactga tataattaaa 18711	
ttgaagctct aatttgtgag tttagtatac atgcatttac ttataataca gttttttagt 18771	
tttgctggcc gcatcttctc aaatatgctt cccagcctgc ttttctgtaa cgttcaccct 18831	
ctaccttage atecettece tttgcaaata gteetettee aacaataata atgteagate 18891	
ctgtagagac cacatcatcc acggttctat actgttgacc caatgcgtct cccttgtcat 18951	

Y.a. (

£12.

79

ctaaacccac accgggtgtc ataatcaacc aatcgtaacc ttcatctctt ccacccatgt 19011 ctctttgagc aataaagccg ataacaaaat ctttgtcgct cttcgcaatg tcaacagtac 19071 ccttagtata ttctccagta gatagggagc ccttgcatga caattctgct aacatcaaaa 19131 ggcctctagg ttcctttgtt acttcttctg ccgcctgctt caaaccgcta acaatacctg 19191 ggcccaccac accgtgtgca ttcgtaatgt ctgcccattc tgctattctg tatacacccg 19251 cagagtactg caatttgact gtattaccaa tgtcagcaaa ttttctgtct tcgaagagta 19311 aaaaattgta.cttggcggat aatgccttta gcggcttaac tgtgccctcc atggaaaaat 19371 cagtcaagat atccacatgt gtttttagta aacaaatttt gggacctaat gcttcaacta 19431 actocagtaa ttoottggtg gtacgaacat ccaatgaagc acacaagttt gtttgctttt 19491 cgtgcatgat attaaatagc ttggcagcaa caggactagg atgagtagca gcacgttcct 19551 tatatgtagc tttcgacatg atttatcttc gtttcctgca ggtttttgtt ctgtgcagtt 19611 gggttaagaa tactgggcaa tttcatgttt cttcaacact acatatgcgt atatatacca 19671 atctaagtct gtgctccttc cttcgttctt ccttctgttc ggagattacc gaatcaaaaa 19731 aatttcaagg aaaccgaaat caaaaaaaag aataaaaaaa aaatgatgaa ttgaaaagct 19791 tatcgat 19798

<210> 11

<211> 1771

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.deltaNS3NS5.pj

<400> 11

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

- Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125
- Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140
- Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys 145 150 155 160
- Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175
- Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val 180 185 190
- Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205
- Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 215 220
- Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235 240
- Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
 245 250 255
- Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe 260 265 270
- Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr 275 280 285
- Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 290 295 300
- Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 305 310 315 320

Quin.

- Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 335
- Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr 340 345 350
- Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp 355 360 365
- Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu 370 380
- Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro 385 390 395 400
- Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val 405 410 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala 420 425 430

- Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu 435 440 445
- Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 450 455 460
- Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 465 470 475 480
- Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 495
- Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
 500 505 510
- Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 515 520 525
- Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn 530 535 540
- Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 545 550 555 560
- Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val 565 570 575
- Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala . 580 585 590
- Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu 595 600 605
- Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 610 620
- Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val 625 630 635 640
- Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val 645 650 655
- Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala 660 670
- Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His 675 680 685
- Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Arg Val 690 695 700
- Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 705 710 715 720
- His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys 740 745 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe

- 755 760 765
- Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile 770 780
- Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys 785 790 795 800
- Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815
- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845
- Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 860
- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875 880
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
1060 1065 1070

- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390

- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 1595 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg 1765 1770

<210> 12

<211> 20220

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core121

<220>

<221> CDS

<222> (12679)..(18354)

<400> 12

atogatocta cocottgogo taaagaagta tatggocta ctaacgottg totttgtote 60 tgtcactaaa cactggatta ttactoccag atacttattt tggactaatt taaatgattt 120 cggatcaacg ttottaatat cgctgaatot toccacaattg atgaaagtag ctaggaagag 180 gaattggtat aaagtttttg tttttgtaaa tottogaagta tactocaaacg aatttagtat 240 tttotcagtg atotoccaga tgctttcacc otcacttaga agtgotttaa gcatttttt 300 actgtggota tttoccttat otgottotto ogatgattog aactgtaatt gcaaactact 360 tacaatatoa gtgatatoag attgatgtt ttgtocatag taaggaataa ttgtaaatto 420 ccaagcagga atcaattot ttaatgaggo ttoccagaatt gctgotttt gcgtottgta 480 tttaaactgg agtgatttat tgacaatato gaaactcago gaattgotta tggtagtatt 540 atagotcatg aatggoot tottgattge tgttocgta tggtgatatoa tocaacataa 600 ataggttagt tocagcagca cottaggtgg tgtttacat gaaggtott caaaccttc 660 cacaaactga cgaacaagca cottaggtgg tgtttacat aataatcaa attgtggoat 720 gcttagooc gatottgtg gcaattgata totagttoa actactotat ttatottgta 780 tottgoagta ttocaacaca ctaacccg aaactcago ctaactcgaa aaactaactt taattgtoct gtttgtottog 840

cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attggtactg 900 eggtatette attteatatt ttaaaaatge acetttgetg etttteetta atttttagae 960 ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtgatc 1020 tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080 agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140 cttatagogt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200 ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260 ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320 gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380 ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agctcgtaag cgtcgttacc caattgctta ctttagttcc accagatcgt caagctgctt 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740 tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaaatg 1860 caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340 ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctg aagacgcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tccacaaaca ttgctcaaaa gtatctcttt gctatatatc tctgtgctat atccctatat 2520 aacctaccca tccacctttc gctccttgaa cttgcatcta aactcgacct ctacatcaac 2580 aggetteeaa tgetetteaa attttaetgt caagtagace cataeggetg taatatgetg 2640

ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820 tattatctcc gcctcagttt gatcttccgc ttcagactgc catttttcac ataatgaatc 2880 tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggcggtagct cctgatctcc 3000 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060 gcacctaata acattettea aggeggagaa gtttgggcca gatgcccaat atgettgaca 3120 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180 tegaagataa gagaagaatg cagtgacett tgtattgaca aatggagatt ccatgtatet 3240 aaaaaatacg cetttaggee ttetgatace ettteeeetg eggtttageg tgeettttae 3300 attaatatet aaaceetete egatggtgge etttaaetga etaataaatg caacegatat 3360 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3480 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540 cccccttgtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3600 cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttattttc 3660 cattacaget gatgteattg tatateageg etgtaaaaat etatetgtta cagaaggttt 3720 tegeggtttt tataaacaaa actttegtta egaaategag caateacece agetgegtat 3780 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840 ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaaac ctcaatctgt 3900 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 3960 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4020 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4080 tetactgtac gatacaette egeteaggte ettgteettt aacgaggeet taccaetett 4140 ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt tttttttt 4320 ttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatctttta 4380 agcaaggatt ttettaaett etteggegae ageateaeeg aetteggtgg taetgttgga 4440

accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4500 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560 agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4620 gtacaaacca aatgcggtgt tcttgtctgg caaagaggcc aaggacgcag atggcaacaa 4680 acceaaggag cetgggataa eggaggette ateggagatg atateaceaa acatgttget 4740 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4800 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4860 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920 tggtggctca tgttgtaggg ccatgaaagc ggccattctt gtgattcttt gcacttctgg 4980 aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5040 aaagtaaata ceteceacta attetetaae aacaacgaag teagtacett tagcaaattg 5100 tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160 ggcgtacaat tgaagttett tacggatttt tagtaaacet tgttcaggte taacactace 5220 ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5340 atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400 aatggcttcg gctgtgattt cttgaccaac gtggtcacct ggcaaaacga cgatcttctt 5460 aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640 acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700 tatagtctag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 5940 gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000 caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6120 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240

taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6300 cacttcccgc gtttactgat tactagcgaa gctgcgggtg cattttttca agataaaggc 6360 atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6540 tottactaca attittitgt otaaagagta atactagaga taaacataaa aaatgtagag 6600 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6660 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720 aatattttag tagetegtta cagteeggtg egtttttggt tttttgaaag tgegtettea 6780 gagegetttt ggtttteaaa agegetetga agtteetata etttetagag aataggaact 6840 teggaatagg aactteaaag egttteegaa aacgageget teegaaaatg caacgegage 6900 tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatatata 6960 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7080 gtategtatg etteetteag eactaceett tagetgttet atatgetgee acteeteaat 7140 tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7200 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260 cctggccacg gcagaagcac gcttatcgct ccaatttccc acaacattag tcaactccgt 7320 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380 attitutata gcaaagatig aataaggcgc attitutctic aaagctitat tgtacgatct 7440 gactaagtta tettttaata attggtatte etgtttattg ettgaagaat tgeeggteet 7500 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7560 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7680 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7740 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7860 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 7980 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8040

ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8160 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8220 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8280 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8400 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520 gtateattge ageactgggg ceagatggta ageceteeeg tategtagtt atetacaega 8580 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8700 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8760 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820 gatettettg agateetttt tttetgegeg taatetgetg ettgeaaaca aaaaaaccae 8880 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 8940 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 9000 accaetteaa gaaetetgta geaeegeeta eataeetege tetgetaate etgttaeeag 9060 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9240 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300 cgagggagct tccaggggga aacgcctggt atctttatag tcctgtcggg tttcgccacc 9360 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9420 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9480 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9600 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9660 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9720 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780 ggettgtetg etceeggeat eegettacag acaagetgtg acegteteeg ggagetgeat 9840

gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9900 agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gctcgttgag 9960 tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080 accgatgaaa cgagagagga tgctcacgat acgggttact gatgatgaac atgcccggtt 10140 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10260 gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10320 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380 tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500 tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10620 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10680 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10740 gegeetacaa tecatgeeaa ecegtteeat gtgetegeeg aggeggeata aategeegtg 10800 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860 tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 10980 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11040 ccgaaacgtt tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11160 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11280 ctcaagggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggtc 11340 ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaatct 11400 atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460 attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcacaa 11520 caataccagt ccggatcaac tggcaccatc tctcccgtag tctcatctaa tttttcttcc 11580 ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640

atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tqtatttcgg gtttgttaac 11700 tegttecaqt caggaatqtt ccaegtgaag ctatetteca qeaaaqtete caettettea 11760 tcaaattgtg gagaatactc ccaatgetet tatetatggg acttccggga aacacagtac 11820 cgatacttcc caattcgtct tcagagctca ttgtttgttt gaagagacta atcaaagaat 11880 cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgtagg 11940 ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000 tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaatcc 12060 tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggctt 12120 tegeteataa aaatgttatg aegttttgee egeaggeggg aaaccateea etteaegaga 12180 ctgatctcct ctgccggaac accgggcatc tccaacttat aagttggaga aataagagaa 12240 tttcagattg agagaatgaa aaaaaaaaac ccttagttca taggtccatt ctcttagcgc 12300 aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgatg 12360 caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgta tctatctcat 12420 tttcttacac cttctattac cttctgctct ctctgatttg gaaaaagctg aaaaaaaagg 12480 ttgaaaccag ttccctgaaa ttattcccct acttgactaa taagtatata aagacggtag 12540 gtattgattg taattctgta aatctatttc ttaaacttct taaattctac ttttatagtt 12600 agtetttttt ttagttttaa aacaccaaga aettagttte gaataaacac acataaacaa 12660 acaagettae aaaacaaa atg get gea tat gea get eag gge tat aag gtg 12711 Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac 12759 Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga 12807 Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg 30 aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc 12855 Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe ctt gcc gac ggc ggg tgc tcg ggg ggc gct tat gac ata ata att tgt 12903 Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc att ggc act 12951 Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr gtc ctt gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc 12999

Val	Leu	Asp	Gln 95	Ala	Glu	Thr	Ala	Gly 100	Ala	Arg	Leu	Val	Val 105	Leu	Ala	
		acc Thr 110														13047
		gct Ala														13095
		ctc Leu														13143
tca Ser	aag Lys	aag Lys	aag Lys	tgc Cys 160	gac Asp	gaa Glu	ctc Leu	gcc Ala	gca Ala 165	aag Lys	ctg Leu	gtc Val	gca Ala	ttg Leu 170	ggc Gly	13191
		gcc Ala														13239
		ggc Gly 190														13287
tat Tyr	acc Thr 205	ggc Gly	gac Asp	ttc Phe	gac Asp	tcg Ser 210	gtg Val	ata Ile	gac Asp	tgc Cys	aat Asn 215	acg Thr	tgt Cys	gtc Val	acc Thr	13335
		gtc Val														13383
acg Thr	ctc Leu	ccc Pro	caa Gln	gat Asp 240	gct Ala	gtc Val	tcc Ser	cgc Arg	act Thr 245	caa Gln	cgt Arg	cgg Arg	ggc Gly	agg Arg 250	act Thr	13431
ggc Gly	agg Arg	gly ggg	aag Lys 255	cca Pro	ggc	atc Ile	tac Tyr	aga Arg 260	ttt Phe	gtg Val	gca Ala	ccg Pro	999 Gly 265	gag Glu	cgc Arg	13479
ccc Pro	tcc Ser	ggc Gly 270	atg Met	ttc Phe	gac Asp	tcg Ser	tcc Ser 275	gtc Val	ctc Leu	tgt Cys	gag Glu	tgc Cys 280	tat Tyr	gac Asp	gca Ala	13527
ggc	tgt Cys 285	gct Ala	tgg Trp	tat Tyr	gag Glu	ctc Leu 290	acg Thr	ccc Pro	gcc Ala	gag Glu	act Thr 295	aca Thr	gtt Val	agg Arg	cta Leu	13575
cga Arg 300	gcg Ala	tac Tyr	atg Met	aac Asn	acc Thr 305	ccg Pro	ggg Gly	ctt Leu	ccc Pro	gtg Val 310	tgc Cys	cag Gln	gac Asp	cat His	ctt Leu 315	13623
gaa Glu	ttt Phe	tgg Trp	gag Glu	ggc Gly 320	gtc Val	ttt Phe	aca Thr	ggc Gly	ctc Leu 325	act Thr	cat His	ata Ile	gat Asp	gcc Ala 330	cac His	13671

	cta L e u															13719
	tac Tyr															13767
	gac Asp 365															13815
	cca Pro			_			_	_		_	_	_		_		13863
	ctg Leu	_			_					_		_	-	_	_	13911
	ctg Leu															13959
	gct Ala															14007
	agg Arg 445															14055
	ctc Leu															14103
	tac Tyr															14151
gcc Ala	ctc Leu	Gly	ctc Leu 495	Leu	Gln	acc Thr	Ala	Ser	Arg	Gln	gca Ala	Glu	Val	Ile	gcc Ala	14199
cct Pro	gct Ala	gtc Val 510	cag Gln	acc Thr	aac Asn	tgg Trp	caa Gln 515	aaa Lys	ctc Leu	gag Glu	acc Thr	ttc Phe 520	tgg Trp	gcg Ala	aag Lys	14247
cat His	atg Met 525	tgg Trp	aac Asn	ttc Phe	atc Ile	agt Ser 530	ggg Gly	ata Ile	caa Gln	tac Tyr	ttg Leu 535	gcg Ala	ggc Gly	ttg Leu	tca Ser	14295
	ctg Leu															14343
gct Ala	gtc Val	acc Thr	agc Ser	cca Pro 560	cta Leu	acc Thr	act Thr	agc Ser	caa Gln 565	acc Thr	ctc Leu	ctc Leu	ttc Phe	aac Asn 570	ata Ile	14391

ttg ggg ggg t Leu Gly Gly 5			u Ala Ala		a Ala Thr	
gcc ttt gtg g Ala Phe Val (590						
ctg ggg aag g Leu Gly Lys v 605						
gcg gga gct o Ala Gly Ala I 620						
acg gag gac o Thr Glu Asp 1						
ctc gta gtc g Leu Val Val (a Ile Leu		s Val Gly	
ccg ggc gag g Pro Gly Glu 6						
tcc cgg ggg a Ser Arg Gly 1 685			r His Tyr			14775
gca gct gcc (Ala Ala Ala A 700						
ctc ctg agg o Leu Leu Arg i						
tgc tcc ggt t Cys Ser Gly s	Ser Trp Leu	Arg Asp Il	e Trp Asp		s Glu Val	14919
ttg agc gac t Leu Ser Asp 1 750						
cct ggg atc o Pro Gly Ile 1 765	ccc ttt gtg Pro Phe Val	tcc tgc ca Ser Cys Gl 770	g cgc ggg n Arg Gly	tat aag gg Tyr Lys Gl 775	g gtc tgg y Val Trp	15015
cga ggg gac o Arg Gly Asp o	ggc atc atg Gly Ile Met 785	cac act cg His Thr Ar	c tgc cac g Cys His 790	tgt gga gc Cys Gly Al	t gag atc a Glu Ile 795	
act gga cat of the Gly His V	gtc aaa aac Val Lys Asn 800	ggg acg at Gly Thr Me	g agg atc t Arg Ile 805	gtc ggt cc Val Gly Pr	t agg acc o Arg Thr 810	15111

tgc agg aac atg t Cys Arg Asn Met T 815			-	_	5159
ggc ccc tgt acc c Gly Pro Cys Thr P 830	ro Leu Pro Al				5207
agg gtg tct gca g Arg Val Ser Ala G 845					5255
cac tac gtg acg g His Tyr Val Thr G 860		hr Asp Asn L			5303
gtc cca tcg ccc g Val Pro Ser Pro G 8					5351
agg ttt gcg ccc c Arg Phe Ala Pro P 895					5399
aga gta gga ctc c Arg Val Gly Leu H 910	is Glu Tyr Pr				5447
ccc gaa ccg gac g Pro Glu Pro Asp V 925					5495
cat ata aca gca g His Ile Thr Ala G 940		ly Arg Arg L			5543
ccc tct gtg gcc a Pro Ser Val Ala S			-		5591
aag gca act tgc a Lys Ala Thr Cys T 975	hr Ala Asn Hi	is Asp Ser P		Leu Ile	5639
gag gcc aac ctc c Glu Ala Asn Leu I 990	eu Trp Arg G				5687
gtt gag tca gaa a Val Glu Ser Glu A 1005					5735
gtg gcg gag gag g Val Ala Glu Glu A 1020			al Pro Ala Glu		5783
cgg aag tct cgg a Arg Lys Ser Arg A	ga ttc gcc ca rg Phe Ala Gl 40	ag gcc ctg c ln Ala Leu P 1045	ro Val Trp Ala	cgg ccg 19 Arg Pro 1050	5831

gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1435	16983
ctc acc gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly 1440 1445 1450	17031
gag aac tgc ggc tat cgc agg tgc cgc gcg agc ggc gta ctg aca act $G\underline{l}u$. Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1455 1460 1465	17079
agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys 1470 1480	17127
cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp 1485 1490 1495	17175
tta gtc gtt atc tgt gaa agc gcg ggg gtc cag gag gac gcg gcg agc Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser 1500 1505 1510 1515	17223
ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc cct ggg Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly 1520 1525 1530	17271

		g gag ctc ata aca u Glu Leu Ile Thr 0 1	
		gc gct gga aag agg y Ala Gly Lys Arg 1560	
		c gcg aga gct gcg eu Ala Arg Ala Ala 1575	
		gg cta ggc aac ata p Leu Gly Asn Ile 1590	
	p Ala Arg Met Il	a ctg atg acc cat e Leu Met Thr His 1605	_
		a cag gcc ctc gat u Gln Ala Leu Asp 0	
	_	ca ctg gat cta cct co Leu Asp Leu Pro 1640	
_		t tca ctc cac agt ne Ser Leu His Ser 1655	
		gc ctc aga aaa ctt vs Leu Arg Lys Leu 1670	
_ -	rp Arg His Arg Al	cc cgg agc gtc cgc a Arg Ser Val Arg 1685	
	ly Arg Ala Ala Il	a tgt ggc aag tac le Cys Gly Lys Tyr 00	Leu Phe Asn
		cc act cca ata gcg eu Thr Pro Ile Ala 1720	
cag ctg gac ttg to Gln Leu Asp Leu So 1725	cc ggc tgg ttc ac er Gly Trp Phe Th 1730	eg get gge tac age nr Ala Gly Tyr Ser 1735	ggg gga gac 17895 Gly Gly Asp
att tat cac agc g Ile Tyr His Ser V 1740	ng tot cat goo co al Ser His Ala An 1745	gg ccc cgc tgg atc gg Pro Arg Trp Ile 1750	tgg ttt tgc 17943 Trp Phe Cys 1755
cta ctc ctg ctt g Leu Leu Leu Leu A 17	la Ala Gly Val Gl	gc atc tac ctc ctc Ly Ile Tyr Leu Leu 1765	ccc aac cga 17991 Pro Asn Arg 1770

atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn 1775 1780 1785	18039
cgg cgg ccg cag gac gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile Val Gly 1790 1795 1800	18087
gga gtt tac ttg ttg ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 1805 1810 1815	18135
acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro 1820 1825 1830 1835	18183
atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 1845 1850	18231
tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg gcg gga tgg Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 1860 1865	18279
ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1875 1880	18327
cgg cgt agg tcg cgc aat ttg ggt aag taatagtcga ctttgttccc Arg Arg Arg Ser Arg Asn Leu Gly Lys 1885 1890	18374
actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta aacaacatg	t 18434
tttcccatgt aatatccttt tctatttttc gttccgttac caactttaca catacttta	t 18494
atagetatte acttetatae actaaaaaac taagacaatt ttaattttge tgeetgeea	t 18554
atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa aagatgaat	g 18614
tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc gccatgatc	g 18674
cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg ccaaagcgg	t 18734
cggacagtgc tccgagaacg ggtgcgcata gaaattgcat caacgcatat agcgctagc	a 18794
gcacgccata gtgactggcg atgctgtcgg aatggacgat atcccgcaag aggcccggc	a 18854
gcacgccata gtgactggcg atgctgtcgg aatggacgat atcccgcaag aggcccggcggtgtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacg	
	a 18914
gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacg	a 18914 a 18974
gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacg	a 18914 a 18974 t 19034
gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacg tgagcgcatt gttagatttc atacacggtg cctgactgcg ttagcaattt aactgtgat aactaccgca ttaaagcttt ttctttccaa ttttttttt ttcgtcatta taaaaatca	a 18914 a 18974 t 19034 t 19094

 ψ_{ik}

ttgcaaatag tcctcttcca acaataataa tgtcagatcc tgtagagacc acatcatcca 19274 eggttetata etgttgacce aatgegtete cettgteate taaacceaca eegggtgtea 19334 taatcaacca atcgtaacct tcatctcttc cacccatgtc tctttgagca ataaagccga 19394 taacaaaatc tttgtcgctc ttcgcaatgt caacagtacc cttagtatat tctccagtag 19454 atagggagee ettgeatgae aattetgeta acateaaaag geetetaggt teetttgtta 19514 cttcttctgc cgcctgcttc aaaccgctaa caatacctgg gcccaccaca ccgtgtgcat 19574 togtaatgtc tgcccattct gctattctgt atacacccgc agagtactgc aatttgactg 19634 tattaccaat gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac ttggcggata 19694 atgcctttag cggcttaact gtgccctcca tggaaaaatc agtcaagata tccacatgtg 19754 tttttagtaa acaaattttg ggacctaatg cttcaactaa ctccagtaat tccttggtgg 19814 tacgaacatc caatgaagca cacaagtttg tttgcttttc gtgcatgata ttaaatagct 19874 tggcagcaac aggactagga tgagtagcag cacgtteett atatgtaget ttegacatga 19934 tttatcttcg tttcctgcag gtttttgttc tgtgcagttg ggttaagaat actgggcaat 19994 ttcatgtttc ttcaacacta catatgcgta tatataccaa tctaagtctg tgctccttcc 20054 ttcgttcttc cttctgttcg gagattaccg aatcaaaaaa atttcaagga aaccgaaatc 20114 aaaaaaaaga ataaaaaaaa aatgatgaat tgaaaagctt atcgat 20160

<210> 13

<211> 1892

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core121

<400> 13

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala

85 90 9**5**

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 \$105\$ 110

- Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125
- Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140
- Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys 145 150 155 160
- Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175
- Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val 180 185 190
- Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205
- Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 215 220
- Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235 240
- Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro 245 250 255
- Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe 260 265 270
- Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr 275 280 285
- Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 290 295 300
- Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 305 310 315 320
- Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 335
- Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr 340 345 350
- Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp 355 360 365
- Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu 370 375 380
- Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro 385 390 395 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val 410 Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala 425 420 Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu 440 Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 455 Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 470 Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 520 Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu 600 605 Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 615 Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His 680 Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val 695 Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 710 715

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp $725 \hspace{1cm} 730 \hspace{1cm} 735$

- Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750
- Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe 755 760 765
- Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile 770 780
- Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys 785 790 795 800
- Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815
- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845
- Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 855
- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875 880
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055

- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360

Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375

- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Glm Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala 745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro 1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp 1780 1785 1790

Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu 1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser 1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Gln Pro Ile Pro Lys Ala Arg 825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu 1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg 1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Ser Arg 1875 1880 1885

Asn Leu Gly Lys 1890

<210> 14

<211> 20316

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core173

<220>

<221> CDS

<222> (12679)..(18510)

<400> 14

ategatecta eccettgege taaagaagta tatgtgeeta etaaegettg tetttgtete 60

tgtcactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120 cggatcaacg ttcttaatat cgctgaatct tccacaattg atgaaagtag ctaggaagag 180 gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240 ttteteagtg ateteceaga tgettteace eteaettaga agtgetttaa geattttttt 300 actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgtaatt gcaaactact 360 tacaatatca gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420 ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgcttttt gcgtcttgta 480 tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540 atageteatg aatgtggete tettgattge tgtteegtta tgtgtaatea teeaacataa 600 ataggttagt teageageae ataatgetat ttteteacet gaaggtettt caaacettte 660 cacaaactga cgaacaagca ccttaggtgg tgttttacat aatatatcaa attgtggcat 720 gettagegee gatettgtgt geaattgata tetagtttea actaetetat ttatettgta 780 tettgeagta tteaaacacg etaactegaa aaactaactt taattgteet gtttgteteg 840 cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attggtactg 900 cggtatette attteatatt ttaaaaatge acetttgetg etttteetta atttttagae 960 ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtgatc 1020 tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080 agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140 cttatagegt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200 ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260 ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320 gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380 ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agctegtaag egtegttace caattgetta etttagttee accagategt caagetgett 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740 tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaaatg 1860

caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340 ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctg aagacgcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tccacaaaca ttgctcaaaa gtatctcttt gctatatatc tctgtgctat atccctatat 2520 aacctaccca tccacctttc gctccttgaa cttgcatcta aactcgacct ctacatcaac 2580 aggetteeaa tgetetteaa attttaetgt caagtagaee cataeggetg taatatgetg 2640 ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcqqc 2820 tattatetee geeteagttt gatetteege tteagaetge cattttteae ataatgaate 2880 tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggcggtagct cctgatctcc 3000 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060 gcacctaata acattettea aggeggagaa gtttgggeca gatgeecaat atgettgaca 3120 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180 tegaagataa gagaagaatg cagtgacett tgtattgaca aatggagatt ccatgtatet 3240 aaaaaatacg cctttaggcc ttctgatacc ctttcccctg cggtttagcg tgccttttac 3300 attaatatet aaaccetete egatggtgge etttaactga etaataaatg caaccgatat 3360 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3480 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540 cccccttgtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3600 cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttattttc 3660

cattacaget gatgteattg tatateageg etgtaaaaat etatetgtta eagaaggttt 3720 tegeggtttt tataaacaaa aetttegtta egaaategag caateaeeee agetgegtat 3780 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840 ttaacttega gaagggatta aggetaattt eactagtatg ttteaaaaac eteaatetgt 3900 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 3960 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4020 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4080 tetactgtae gatacaette egeteaggte ettgteettt aacgaggeet taccaetett 4140 ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt tttttttt 4320 ttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatctttta 4380 agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcggtgg tactgttgga 4440 accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4500 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560 agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4620 gtacaaacca aatgeggtgt tettgtetgg caaagaggee aaggaegeag atggeaacaa 4680 acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgct 4740 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4800 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4860 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920 tggtggetea tgttgtaggg ccatgaaagc ggccattett gtgattettt gcaettetgg 4980 aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5040 aaagtaaata cctcccacta attctctaac aacaacgaag tcagtacctt tagcaaattg 5100 tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160 ggcgtacaat tgaagttett tacggatttt tagtaaacct tgttcaggte taacactace 5220 ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5340 atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400 aatggcttcg gctgtgattt cttgaccaac gtggtcacct ggcaaaacga cgatcttctt 5460

aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640 acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700 tataqtctaq cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 5940 gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000 caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6120 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240 taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6300 cacttcccgc gtttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6360 atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6540 tettactaca attttttgt etaaagagta atactagaga taaacataaa aaatgtagag 6600 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6660 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720 aatattttag tagctcgtta cagtccggtg cgtttttggt tttttgaaag tgcgtcttca 6780 gagegetttt ggttttcaaa agegetetga agtteetata etttetagag aataggaact 6840 toggaatagg aacttcaaag ogtttoogaa aacgagogot toogaaaatg caacgogago 6900 tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatatata 6960 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7080 gtatcgtatg cttccttcag cactaccctt tagctgttct atatgctgcc actcctcaat 7140 tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7200 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260

cetggccacg gcagaagcac gettateget ccaattteec acaacattag tcaacteegt 7320 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380 attitttata gcaaagattg aataaggcgc attittcttc aaagctitat tgtacgatct 7440 gactaagtta tettttaata attggtatte etgtttattg ettgaagaat tgeeggteet 7500 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7560 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7680 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7740 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7860 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920 tgggttacat cgaactggat ctcaacagcg gtaagatect tgagagtttt cgccccgaag 7980 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8040 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8160 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8220 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8280 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8400 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520 gtateattge ageactgggg ccagatggta ageceteeeg tategtagtt atetacaega 8580 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8700 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8760 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8880 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 8940 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 9000 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 9060

tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9240 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300 cgagggaget tecaggggga aacgeetggt atetttatag teetgteggg tittegeeace 9360 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9420 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9480 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9600 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9660 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9720 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780 ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 9840 gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9900 agegtggteg tgaagegatt cacagatgte tgeetgttea teegegteea getegttgag 9960 tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080 accgatgaaa cgagagagga tgctcacgat acgggttact gatgatgaac atgcccggtt 10140 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10260 gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10320 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380 tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500 tggccaggac ccaacgctgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10620 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10680 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10740 gegeetacaa teeatgeeaa eeegtteeat gtgetegeeg aggeggeata aategeegtg 10800 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860

tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 10980 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11040 ccgaaacgtt tggtggcggg accagtgacg aaggettgag cgagggcgtg caagattecg 11100 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11160 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11280 ctcaagggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggtc 11340 ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaatct 11400 atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460 attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcacaa 11520 caataccagt ccggatcaac tggcaccatc tctcccgtag tctcatctaa tttttcttcc 11580 ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataga 11640 atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatttcgg gtttgttaac 11700 tegttecagt caggaatgtt ccacgtgaag ctatetteca geaaagtete caettettea 11760 tcaaattgtg gagaatactc ccaatgctct tatctatggg acttccggga aacacagtac 11820 cgatacttcc caattcgtct tcagagctca ttgtttgttt gaagagacta atcaaagaat 11880 cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgtagg 11940 ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000 tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaatcc 12060 tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggctt 12120 tegeteataa aaatgttatg aegttttgee egeaggeggg aaaceateea etteaegaga 12180 ctgatctcct ctgccggaac accgggcatc tccaacttat aagttggaga aataagagaa 12240 tttcagattg agagaatgaa aaaaaaaac ccttagttca taggtccatt ctcttagcgc 12300 aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtgatg 12360 caacctgcct ggagtaaatg atgacacaag gcaattgacc cacgcatgta tctatctcat 12420 tttcttacac cttctattac cttctgctct ctctgatttg gaaaaagctg aaaaaaagg 12480 ttgaaaccag ttccctgaaa ttattcccct acttgactaa taagtatata aagacggtag 12540 gtattgattg taattetgta aatetattte ttaaaettet taaattetae ttttatagtt 12600 agtetttttt ttagttttaa aacaccaaga aettagttte gaataaacac acataaacaa 12660

acaagcttac aaaacaaa			Gln Gly Tyr Ly	
cta gta ctc aac ccc Leu Val Leu Asn Pro 15				
atg tcc aag gct cat Met Ser Lys Ala His 30				
aca att acc act ggc Thr Ile Thr Thr Gly 45				
ctt gcc gac ggc ggg Leu Ala Asp Gly Gly 60			-	_
gac gag tgc cac tcc Asp Glu Cys His Ser 80			Leu Gly Ile G	
gtc ctt gac caa gca Val Leu Asp Gln Ala 95	Glu Thr Ala			
acc gcc acc cct ccg Thr Ala Thr Pro Pro 110				
gag gtt gct ctg tcc Glu Val Ala Leu Ser 125		Glu Ile Pro		
atc ccc ctc gaa gta Ile Pro Leu Glu Val 140			-	•
tca aag aag aag tgc Ser Lys Lys Lys Cys 160	Asp Glu Leu		Leu Val Ala Le	
atc aat gcc gtg gcc Ile Asn Ala Val Ala 175	tac tac cgc Tyr Tyr Arg	ggt ctt gac Gly Leu Asp 180	gtg tcc gtc at Val Ser Val II 185	cc ccg 13239 Le Pro
acc agc ggc gat gtt Thr Ser Gly Asp Val 190	gtc gtc gtg Val Val Val 195	gca acc gat Ala Thr Asp	gcc ctc atg ac Ala Leu Met Tl 200	ec ggc 13287 nr Gly
tat acc ggc gac ttc Tyr Thr Gly Asp Phe 205	gac tcg gtg Asp Ser Val 210	Ile Asp Cys .	aat acg tgt g Asn Thr Cys Va 215	c acc 13335 al Thr
cag aca gtc gat ttc Gln Thr Val Asp Phe 220	agc ctt gac Ser Leu Asp 225	cct acc ttc Pro Thr Phe 230	acc att gag ac Thr Ile Glu Tl	ea atc 13383 ur Ile 235

					gct Ala											13431
					ggc Gly											13479
					gac Asp											13527
	_	-			gag Glu		_		_				_			13575
					acc Thr 305											13623
_					gtc Val								_	_		13671
					aag Lys											13719
			_		gtg Val	_	_		_		_				_	13767
					aag Lys											13815
					cta Leu 385											13863
					gtc Val		Lys	Tyr	Ile	Met		Cys		Ser		13911
gac Asp	ctg Leu	gag Glu	gtc Val 415	gtc Val	acg Thr	agc Ser	acc Thr	tgg Trp 420	gtg Val	ctc Leu	gtt Val	ggc Gly	ggc Gly 425	gtc Val	ctg Leu	13959
					tat Tyr											14007
ggc Gly	agg Arg 445	gtc Val	gtc Val	ttg Leu	tcc Ser	999 Gly 450	aag Lys	ccg Pro	gca Ala	atc Ile	ata Ile 455	cct Pro	gac Asp	agg Arg	gaa Glu	14055
gtc Val 460	ctc Leu	tac Tyr	cga Arg	gag Glu	ttc Phe 465	gat Asp	gag Glu	atg Met	gaa Glu	gag Glu 470	tgc Cys	tct Ser	cag Gln	cac His	tta Leu 475	14103

ccg tac Pro Tyr			_	_	Leu l	_		_		_	_	_	14151
gcc ctc Ala Leu				Ala S									14199
cct gct Pro Ala			Trp (_	14247
cat atg His Met 525									_		_		14295
acg ctg Thr Leu 540			Āla	_	-		_	_	-			-	14343
gct gtc Ala Val	_				Ser (14391
ttg ggg Leu Gly				Gln I									14439
gcc ttt Ala Phe			Leu				-			_	_		14487
		ctc ata Leu Ile											14535
		gtg gca Val Ala 625	Phe										14583
	Asp Leu	gtc aat Val Asn 640	Leu	Leu I	Pro	Ãla	Ile		_	Pro		Ala	14631
		gtg gtc Val Val		Ala A									14679
		gca gtg Ala Val	Gln										14727
		cat gtt His Val			_			-	_		_	_	14775
		gtc act Val Thr 705	Ala										14823

					cac His											14871
					cta Leu											14919
_	_	_		_	acc Thr				_	_		_		_	_	14967
					gtg Val											15015
					atg Met 785											15063
			_		aac Asn		_	_			_					15111
					agt Ser							_			_	15159
		_			ctt Leu			_			-					15207
			_		gaa Glu						-			_		15255
					atg Met 865											15303
					ttt Phe	Phe	Thr	Glu	Leu	Asp		Val	Arg			15351
					tgc Cys											15399
aga Arg	gta Val	gga Gly 910	ctc Leu	cac His	gaa Glu	tac Tyr	ccg Pro 915	Val	gly ggg	tcg Ser	caa Gln	tta Leu 920	cct Pro	tgc Cys	gag Glu	15447
ccc Pro	gaa Glu 925	ccg Pro	gac Asp	gtg Val	gcc Ala	gtg Val 930	ttg Leu	acg Thr	tcc Ser	atg Met	ctc Leu 935	act Thr	gat Asp	ccc Pro	tcc Ser	15495
					gcg Ala 945											15543

t. 12

ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	15591
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	15639
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	15687
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	15735
gtg gcg gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctg Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020 1025 1030 1035	15783
cgg aag tct cgg aga ttc gcc cag gcc ctg ccc gtt tgg gcg cgg ccg Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040 1045 1050	15831
gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gaa Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	15879
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1075 1080	15927
gtg cct ccg cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	15975
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1115	16023
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro 1120 1125 1130	16071
gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263

gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791
gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala 1375 1380 1385	16839
tgg aag tcc aag aaa acc cca atg ggg ttc tcg tat gat acc cgc tgc Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1390 1395 1400	16887
ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca atc Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile 1405 1410 1415	16935
tac caa tgt tgt gac ctc gac ccc caa gcc cgc gtg gcc atc aag tcc Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser 1420 1425 1430 1435	16983

ctc acc gag agg ctt Leu Thr Glu Arg Leu 1440	Tyr Val Gly G			17031
gag aac tgc ggc tat Glu Asn Cys Gly Tyr 1455		rg Ala Ser Gly		17079
age tgt ggt aac acc Ser Cys Gly Asn Thi 1470		yr Ile Lys Ala		17127
cga gcc gca ggg ctc Arg Ala Ala Gly Lec 1485				17175
tta gtc gtt atc tgt Leu Val Val Ile Cys 1500			-	17223
ctg aga gcc ttc acc Leu Arg Ala Phe Thi 1520	Glu Ala Met T			17271
gac ccc cca caa cca Asp Pro Pro Gln Pro 1535	Glu Tyr Asp L		_	17319
tcc aac gtg tca gtc Ser Asn Val Ser Val 1550		ly Ala Gly Lys		17367
ctc acc cgt gac cct Leu Thr Arg Asp Pro 1565				17415
gca aga cac act cca Ala Arg His Thr Pro 1580	_		_	17463
gcc ccc aca ctg tgg Ala Pro Thr Leu Trp 1600		le Leu Met Thr		17511
gtc ctt ata gcc agg Val Leu Ile Ala Arg 1615	g Asp Gln Leu G			17559
tac ggg gcc tgc tac Tyr Gly Ala Cys Ty: 1630		ro Leu Asp Leu		17607
caa aga ctc cat ggo Gln Arg Leu His Gly 1645				17655
ggt gaa atc aat agg Gly Glu Ile Asn Arg 1660				17703

£...

ccc ttg cga gct tgg ag Pro Leu Arg Ala Trp Ar 1680	g His Arg Ala A			17751
ctg gcc aga gga ggc ag Leu Ala Arg Gly Gly Ar 1695		ys Gly Lys Tyr		17799
tgg gca gta aga aca aa Trp Ala Val Arg Thr Ly 1710				17847
cag ctg gac ttg tcc gg Gln Leu Asp Leu Ser Gl 1725				17895
att tat cac agc gtg to Ile Tyr His Ser Val Se 1740 174	r His Ala Arg P			17943
cta ctc ctg ctt gct gc Leu Leu Leu Leu Ala Al 1760	a Gly Val Gly I		_	17991
atg agc acg aat cct as Met Ser Thr Asn Pro Ly 1775		ys Thr Lys Arg		18039
cgg cgg ccg cag gac gt Arg Arg Pro Gln Asp Va 1790				18087
gga gtt tac ttg ttg co Gly Val Tyr Leu Leu Pr 1805				18135
acg aga aag act tcc ga Thr Arg Lys Thr Ser Gl 1820 182	u Arg Ser Gln P			18183
atc ccc aag gct cgt cg Ile Pro Lys Ala Arg Ar 1840		rg Thr Trp Ala		18231
tac cct tgg ccc ctc ta Tyr Pro Trp Pro Leu Ty 1855		ly Cys Gly Trp		18279
ctc ctg tct ccc cgt gg Leu Leu Ser Pro Arg Gl 1870				18327
cgg cgt agg tcg cgc aa Arg Arg Arg Ser Arg As 1885				18375
ggc ttc gcc gac ctc at Gly Phe Ala Asp Leu Me 1900 190	et Gly Tyr Ile P			18423

gga ggc gct gcc agg gcc ctg gcg cat ggc gtc cgg gtt ctg gaa gac 18471 Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp 1920 1925 1930

ggc gtg aac tat gca aca ggg aac ctt cct ggt tgc tct taatagtcga 18520 Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser 1935 1940

ctttqttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 18580 aacaacatgt tttcccatgt aatatccttt tctatttttc gttccgttac caactttaca 18640 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18700 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18760 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18820 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18880 ccaaagcggt cggacagtgc tccgagaacg ggtgcgcata gaaattgcat caacgcatat 18940 agegetagea geaegeeata gtgactggeg atgetgtegg aatggaegat atccegeaag 19000 aggeceggea gtaceggeat aaccaageet atgeetacag catecagggt gaeggtgeeg 19060 aggatgacga tgagcgcatt gttagatttc atacacggtg cctgactgcg ttagcaattt 19120 taaaaaatcat tacgaccgag attcccgggt aataactgat ataattaaat tgaaqctcta 19240 atttgtgagt ttagtataca tgcatttact tataatacag ttttttagtt ttgctggccg 19300 catcttctca aatatgcttc ccagcctgct tttctgtaac gttcaccctc taccttagca 19360 tecetteeet tigeaaatag teetetteea acaataataa tgteagatee tgtagagaee 19420acatcatcca cggttctata ctgttgaccc aatgcgtctc ccttgtcatc taaacccaca 19480 ccgggtgtca taatcaacca atcgtaacct tcatctcttc cacccatqtc tctttqaqca 19540 ataaagccga taacaaaatc tttgtcgctc ttcgcaatgt caacagtacc cttagtatat 19600 tetecagtag atagggagee ettgeatgae aattetgeta acateaaaag geetetaggt 19660 tectttgtta ettettetge egeetgette aaacegetaa caatacetgg geecaceaca 19720 ccgtgtgcat tcgtaatgtc tgcccattct gctattctgt atacacccgc agagtactgc 19780 aatttgactg tattaccaat gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac 19840 ttggcggata atgcctttag cggcttaact gtgccctcca tggaaaaatc agtcaagata 19900 tccacatgtg tttttagtaa acaaattttg ggacctaatg cttcaactaa ctccagtaat 19960 tccttggtgg tacgaacatc caatgaagca cacaagtttg tttgcttttc gtgcatgata 20020 ttaaataget tggcagcaac aggactagga tgagtagcag cacgtteett atatgtaget 20080

ttcgacatga tttatcttcg tttcctgcag gtttttgttc tgtgcagttg ggttaagaat 20140
actgggcaat ttcatgtttc ttcaacacta catatgcgta tatataccaa tctaagtctg 20200
tgctccttcc ttcgttcttc cttctgttcg gagattaccg aatcaaaaaa atttcaagga 20260
aaccgaaatc aaaaaaaaga ataaaaaaaa aatgatgaat tgaaaagctt atcgat 20316

<210> 15

<211> 1944

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core173

<400> 15

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly 35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125.

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys 145 150 155 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val 180 185 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205

Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe

0

210 215 220 Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 230 235 Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro 250 Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 310 Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr 345 Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp 360 Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr 505

Ç.,

Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe

Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
530 540

- Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 545 550 555 560
- Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val 565 570 575
- Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala 580 585 590
- Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu 595 600 605
- Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 610 620
- Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val 625 630 635 640
- Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val 645 650 655
- Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala 660 665 670
- Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His 675 680 685
- Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Arg Val 690 695 700
- Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 705 710 715 720
- His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735
- Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys 740 745 750
- Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
 755 760 765
- Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile 770 780
- Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys 785 790 795 800
- Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815
- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 855 860

- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875 888
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val 1155 1160 1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1180

- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500

- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1735 1740
- Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala 745 1750 1755 1760
- Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro 1765 1770 1775
- Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp 1780 1785 1790
- Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu 1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser 1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Gln Pro Ile Pro Lys Ala Arg 825 1830 1835 1840

Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu 1845 1850 1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg 1860 1865 1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg 1875 1880 1885

Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu 1890 1895 1900

Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg 905 1910 1915 1920

Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala 1925 1930 1935

Thr Gly Asn Leu Pro Gly Cys Ser 1940

<210> 16

<211> 20217

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core140

<220>

<221> CDS

<222> (12679)..(18411)

<400> 16

atcgatccta ccccttgcgc taaagaagta tatgtgccta ctaacgcttg tctttgtctc 60 tgtcactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120 cggatcaacg ttcttaatat cgctgaatct tccacaattg atgaaagtag ctaggaagag 180 gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240 tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300 actgtggcta tttcccttat ctgcttctc cgatgattcg aactgtaatt gcaaactact 360 tacaatatca gtgatatcag attgatgtt ttgtccatag taaggaataa ttgtaaattc 420 ccaagcagga atcaattct ttaatgaggc ttccagaatt gttgcttttt gcgtcttgta 480 tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540

atageteatg aatgtggete tettgattge tgtteegtta tgtgtaatea tecaacataa 600 ataggttagt tcagcagcac ataatgctat tttctcacct gaaggtcttt caaacctttc 660 cacaaactga cgaacaagca ccttaggtgg tgttttacat aatatatcaa attgtggcat 720 gcttagegcc gatcttgtgt gcaattgata tctagtttca actactctat ttatcttgta 780 tcttgcagta ttcaaacacg ctaactcgaa aaactaactt taattgtcct gtttgtctcg 840 cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attggtactg 900 cggtatcttc atttcatatt ttaaaaatgc acctttgctg cttttcctta atttttagac 960 ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtgatc 1020 tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080 agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140 cttatagcgt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200 ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260 ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320 gcgtatette tgactecagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380 ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agctcgtaag cgtcgttacc caattgctta ctttagttcc accagatcgt caagctgctt 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740 tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgctaa tacaaatgaa gaagttccct ccaggcgaac aagaagtaga gacacaaatg 1860 caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340

ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctg aagacgcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tccacaaaca ttgctcaaaa gtatctcttt gctatatatc tctgtgctat atccctatat 2520 aacctaccca tccacctttc gctccttgaa cttgcatcta aactcgacct ctacatcaac 2580 aggettecaa tgetetteaa attttaetgt caagtagace cataeggetg taatatgetg 2640 ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820 tattatetee geeteagttt gatetteege tteagaetge cattttteae ataatgaate 2880 tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggcggtagct cctgatctcc 3000 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060 gcacctaata acattettea aggeggagaa gtttgggcca gatgcccaat atgettgaca 3120 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180 tcgaagataa gagaagaatg cagtgacctt tgtattgaca aatggagatt ccatgtatct 3240 aaaaaatacg cctttaggcc ttctgatacc ctttcccctg cggtttagcg tgccttttac 3300 attaatatct aaaccctctc cgatggtggc ctttaactga ctaataaatg caaccqatat 3360 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3480 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540 cccccttgtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3600 cataatacat tgcttaatac aagcaagcag tetetegeca tteatattte agttatttte 3660 cattacagct gatgtcattg tatatcagcg ctgtaaaaat ctatctgtta cagaaggttt 3720 tegeggtttt tataaacaaa actttegtta egaaategag caateacece agetgegtat 3780 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840 ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaac ctcaatctgt 3900 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 3960 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4020 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4080 tctactgtac gatacacttc cgctcaggtc cttgtccttt aacgaggcct taccactctt 4140

ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt tttttttt 4320 tttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatctttta 4380 agcaaggatt ttettaaett etteggegae agcateaeeg aetteggtgg taetgttgga 4440 accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4500 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560 agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4620 gtacaaacca aatgcggtgt tcttgtctgg caaagaggcc aaggacgcag atggcaacaa 4680 acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgct 4740 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4800 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4860 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920 tggtggctca tgttgtaggg ccatgaaagc ggccattctt gtgattcttt gcacttctgg 4980 aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5040 aaagtaaata cctcccacta attctctaac aacaacgaag tcagtacctt tagcaaattg 5100 tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160 ggcgtacaat tgaagttctt tacggatttt tagtaaacct tgttcaggtc taacactacc 5220 ggtaccccat ttaggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5340 atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taaqaacctt 5400 aatggcttcg gctgtgattt cttgaccaac gtggtcacct ggcaaaacga cgatcttctt 5460 aaaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640 acatecgaac etgggagttt teeetgaaac agatagtata tttgaacetg tataataata 5700 tatagtctag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 5940

gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000 caaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttg 6120 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240 taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6300 cacttcccgc gtttactgat tactagcgaa gctgcgggtg cattttttca agataaaggc 6360 atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca qaaaqtqata 6420 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6540 tcttactaca attttttgt ctaaagagta atactagaga taaacataaa aaatgtagag 6600 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6660 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720 aatattttag tagctcgtta cagtccggtg cgtttttggt tttttgaaag tgcgtcttca 6780 gagegetttt ggtttteaaa agegetetga agtteetata etttetagag aataggaact 6840 teggaatagg aactteaaag egttteegaa aacgageget teegaaaatg caacgegage 6900 tgcgcacata cagctcactg ttcacgtcgc acctatatct gcgtgttgcc tgtatatata 6960 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7080 gtategtatg etteetteag cactaceett tagetgttet atatgetgee acteeteaat 7140 tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7200 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260 cctggccacg gcagaagcac gcttatcgct ccaatttccc acaacattag tcaactccgt 7320 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380 attititata gcaaagattg aataaggcgc attititctic aaagctitat tgtacgatct 7440 gactaagtta tettttaata attggtatte etgtttattg ettgaagaat tgeeggteet 7500 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7560 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7680 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7740

cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7860 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 7980 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8040 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8160 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8220 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8280 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8400 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacqa 8580 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8700 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8760 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 8880 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 8940. ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 9000 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 9060 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9240 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300 cgagggagct tccaggggga aacgcctggt atctttatag tcctgtcggg tttcgccacc 9360 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 9420 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9480 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540

ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9600 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9660 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9720 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780 ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 9840 gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9900 agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gctcgttgag 9960 tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080 accgatgaaa cgagagagga tgctcacgat acgggttact gatgatgaac atgcccggtt 10140 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10260 gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10320 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380 tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500 tggccaggac ccaacgetge eegagatgeg eegegtgegg etgetggaga tggeggaege 10560 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10620 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10680 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10740 gegeetacaa teeatgeeaa eeegtteeat gtgetegeeg aggeggeata aategeegtg 10800 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860 tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 10980 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11040 ccgaaacgtt tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccq 11100 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11160 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11280 ctcaagggca tcggtcgagg atccttcaat atgcgcacat acgctgttat gttcaaggtc 11340

ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaat	ct 11400
atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgt	ta 11460
attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcac	aa 11520
caataccagt coggatcaac tggcaccatc tctcccgtag tctcatctaa tttttctt	cc 11580
ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataata	ga 11640
atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatttcgg gtttgtta	ac 11700
tcgttccagt caggaatgtt ccacgtgaag ctatcttcca gcaaagtctc cacttctt	ca 11760
tcaaattgtg gagaatactc ccaatgctct tatctatggg acttccggga aacacagt	ac 11820
cgatacttcc caattcgtct tcagagctca ttgtttgttt gaagagacta atcaaaga	at 11880
cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgta	gg 11940
ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagt	ct 12000
tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaat	cc 12060
tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggc	tt 12120
tegeteataa aaatgttatg aegttttgee egeaggeggg aaaccateea etteaega	ga 12180
ctgatctcct ctgccggaac accgggcatc tccaacttat aagttggaga aataagag	aa 12240
tttcagattg agagaatgaa aaaaaaaac ccttagttca taggtccatt ctcttagc	gc 12300
aactacagag aacaggggca caaacaggca aaaaacgggc acaacctcaa tggagtga	tg 12360
caacetgeet ggagtaaatg atgacacaag gcaattgace cacgcatgta tetatete	at 12420
tttcttacac cttctattac cttctgctct ctctgatttg gaaaaagctg aaaaaaaa	gg 12480
ttgaaaccag ttccctgaaa ttattcccct acttgactaa taagtatata aagacggt	ag 12540
gtattgattg taattctgta aatctatttc ttaaacttct taaattctac ttttatag	tt 12600
agtettttt ttagttttaa aacaccaaga acttagttte gaataaacac acataaac	aa 12660
acaagcttac aaaacaaa atg gct gca tat gca gct cag ggc tat aag gtg Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val 1 5 10	
cta gta ctc aac ccc tct gtt gct gca aca ctg ggc ttt ggt gct tac Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr 15 20 25	12759
atg tcc aag gct cat ggg atc gat cct aac atc agg acc ggg gtg aga Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg 30 35 40	
aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe 45 50 55	12855

					tgc Cys 65										12903
_		_			acg Thr	_	_				_				12951
					gag Glu										12999
					ggc Gly										13047
					acc Thr										13095
			_	_	atc Ile 145	_			_					_	13143
	_	-	_	_	gac Asp	_		_	_	_	_	~	_	_	 13191
					tac Tyr										13239
					gtc Val										13287
					gac Asp										13335
_		_	-		agc Ser 225	Leu	_	Pro	Thr		Thr				13383
					gct Ala										13431
					ggc										13479
					gac Asp										13527
					gag Glu										13575

cga gcg Arg Ala 300	-	_			_					_	_	_			13623
gaa tti Glu Phe															13671
ttt cta Phe Lei		_		_	_								_	_	13719
gcg tac Ala Ty		_			_	_		_		-				-	13767
tgg gad Trp Asp 36!	Gln														13815
ggg cca Gly Pro 380			_			_	_		_	_	_		_		13863
acc cto															13911
gac cto		_	_	-	_					_			_	-	13959
get ge Ala Ala	-	-								_	-				14007
ggc agg Gly Arg 449	y Val	_	_			_	_	-				_		_	14055
gtc ctc Val Le 460		Arg	Glu	Phe	Asp	Glu	_	Glu	Glu	Сув	Ser	-	His		14103
ccg ta Pro Ty	c atc r Ile	gag Glu	caa Gln 480	Gly 999	atg Met	atg Met	ctc Leu	gcc Ala 485	gag Glu	cag Gln	ttc Phe	aag Lys	cag Gln 490	aag Lys	14151
gcc ct Ala Le															14199
cct gc Pro Al		Gln													14247
cat ate His Me 52	t Trp														14295

	ggt aac ccc Gly Asn Pro 545	-		_	_	14343
	agc cca cta Ser Pro Leu 560	_		Leu Phe		14391
	tgg gtg gct Trp Val Ala 575					14439
	ggc gct ggc Gly Ala Gly					14487
	gtc ctc ata Val Leu Ile			Gly Ala		14535
	ctt gtg gca Leu Val Ala 625	Phe Lys Ile				14583
	ctg gtc aat Leu Val Asn 640	_	_	Ser Pro		14631
	ggc gtg gtc Gly Val Val 655		Ile Leu Arg			14679
	g ggg gca gtg n Gly Ala Val)			_	_	14727
	, aac cat gtt , Asn His Val			Pro Glu		14775
	e cgc gtc act Arg Val Thr 705	Ala Ile Leu				14823
	g cga ctg cac g Arg Leu His 720			Cys Thr		14871
	tcc tgg cta Ser Trp Leu 735		Trp Asp Trp			14919
	ttt aag acc Phe Lys Thr					14967
cct ggg ato Pro Gly Ile 765	ccc ttt gtg Pro Phe Val	tcc tgc cag Ser Cys Gln 770	cgc ggg tat Arg Gly Tyr 775	Lys Gly	gtc tgg Val Trp	15015

cga ggg gac gg Arg Gly Asp Gl 780				Ala Glu I	
act gga cat gt Thr Gly His Va			Ile Val Gly		
tgc agg aac at Cys Arg Asn Me 81	et Trp Ser Gly				
ggc ccc tgt ac Gly Pro Cys Th 830			_		
agg gtg tct go Arg Val Ser Al 845		Val Glu Ile			
cac tac gtg ac His Tyr Val Th 860		_	_	Pro Cys C	_
gtc cca tcg co Val Pro Ser Pi	-		Asp Gly Val	_	
agg ttt gcg co Arg Phe Ala Pi 89					
aga gta gga ct Arg Val Gly Le 910	tc cac gaa tac eu His Glu Tyr		_		
ccc gaa ccg ga Pro Glu Pro Aa 925	ac gtg gcc gtg sp Val Ala Val 930	Leu Thr Ser			
cat ata aca go His Ile Thr A 940	0.45		Leu Ala Arg	Gly Ser I	
	cc agc tcc tcg la Ser Ser Ser 960		Leu Ser Ala		
Lys Ala Thr C	gc acc gct aac ys Thr Ala Asr 75				
gag gcc aac c Glu Ala Asn Lo 990	tc cta tgg agg eu Leu Trp Arg	g cag gag atg g Gln Glu Met 995	ggc ggc aac Gly Gly Asn 1000	atc acc a	agg 15687 Arg
	aa aac aaa gtg lu Asn Lys Val 1010	. Val Ile Leu	_	_	

gtg gcg Val Ala 1020			Asp					Ser					Ile		15783
cgg aag Arg Lys		Arg					Āla					Ala			15831
gac tat Asp Tyr	Asn		_			Glu	_			_	Pro	_		_	15879
cca cct Pro Pro		_			Сув	_				Pro	_				15927
gtg cct Val Pro 1085	Pro			Lys			_		Val			_			15975
cta to Leu Ser 1100			Leu					Thr					Ser		16023
tca act		Gly					Asn					Ser			16071
gcc cct Ala Pro	Ser		_			Āsp		_	_		Ser				16119
atg cco Met Pro					Glu					Asp					16167
tca tgg Ser Trp 1169	Ser			Ser					Ala						16215
tgc tca Cys Se: 1180	_		Tyr		Trp			Āla		_		_	Cys	_	16263
gcg gaa Ala Gl		Gln					Asn					Ser			16311
cgt cac Arg His	s His		-			Ser				_	Ser	_	_		16359
agg cag Arg Gl					Phe					Val					16407
tac cas Tyr Gla 124	qaA n			Lys					Ala						16455

gct aac ttg cta tcc gta Ala Asn Leu Leu Ser Val 1260 1265	Glu Glu Ala Cys S		503
tca gcc aaa tcc aag ttt Ser Ala Lys Ser Lys Phe 1280			551
gcc aga aag gcc gta acc Ala Arg Lys Ala Val Thr 1295			599
gaa gac aat gta aca cca Glu Asp Asn Val Thr Pro 1310			647
gtt ttc tgc gtt cag cct Val Phe Cys Val Gln Pro 1325			695
atc gtg ttc ccc gat ctg Ile Val Phe Pro Asp Leu 1340 1345	Gly Val Arg Val		743
tac gac gtg gtt aca aag Tyr Asp Val Val Thr Lys 1360			791
gga ttc caa tac tca cca Gly Phe Gln Tyr Ser Pro 1375			839
tgg aag tcc aag aaa acc Trp Lys Ser Lys Lys Thr 1390		-	887
ttt gac tcc aca gtc act Phe Asp Ser Thr Val Thr 1405			935
tac caa tgt tgt gac ctc Tyr Gln Cys Cys Asp Leu 1420 1425	Asp Pro Gln Ala	Arg Val Ala Ile Lys Ser	983
ctc acc gag agg ctt tat Leu Thr Glu Arg Leu Tyr 1440			031
gag aac tgc ggc tat cgc Glu Asn Cys Gly Tyr Arg 1455			079
agc tgt ggt aac acc ctc Ser Cys Gly Asn Thr Leu 1470			127
cga gcc gca ggg ctc cag Arg Ala Ala Gly Leu Gln 1485			175

. . .

tta gtc gtt atc tgt gaa Leu Val Val Ile Cys Glu 1500 1505	Ser Ala Gly Val Gl	n Glu Asp Ala Ala Ser
ctg aga gcc ttc acg gag Leu Arg Ala Phe Thr Glu 1520		
gac ccc cca caa cca gaa Asp Pro Pro Gln Pro Glu 1535		
tcc aac gtg tca gtc gcc Ser Asn Val Ser Val Ala 1550		
ctc acc cgt gac cct aca Leu Thr Arg Asp Pro Thr 1565		
gca aga cac act cca gto Ala Arg His Thr Pro Val 1580 1585	Asn Ser Trp Leu Gl	y Asn Ile Ile Met Phe
gcc ccc aca ctg tgg gcg Ala Pro Thr Leu Trp Ala 1600		
gtc ctt ata gcc agg gac Val Leu Ile Ala Arg Asp 1615		
tac ggg gcc tgc tac tcc Tyr Gly Ala Cys Tyr Ser 1630		
caa aga ctc cat ggc ctc Gln Arg Leu His Gly Leu 1645		_
ggt gaa atc aat agg gtg Gly Glu Ile Asn Arg Val 1660 1665	Ala Ala Cys Leu Ar	g Lys Leu Gly Val Pro
ccc ttg cga gct tgg aga Pro Leu Arg Ala Trp Arg 1680	cac cgg gcc cgg ag His Arg Ala Arg Se 1685	c gtc cgc gct agg ctt 1775 r Val Arg Ala Arg Leu 1690
ctg gcc aga gga ggc agg Leu Ala Arg Gly Gly Arg 1695	gct gcc ata tgt gg Ala Ala Ile Cys Gl 1700	c aag tac ctc ttc aac 1779 y Lys Tyr Leu Phe Asn 1705
tgg gca gta aga aca aag Trp Ala Val Arg Thr Lys 1710		
cag ctg gac ttg tcc ggc Gln Leu Asp Leu Ser Gly 1725	tgg ttc acg gct gg Trp Phe Thr Ala Gl 1730	c tac agc ggg gga gac 1789 y Tyr Ser Gly Gly Asp 1735

cta ctc ctg ctt gct gca ggg gta ggc atc tac ctc ctc ccc aac cga l7991 Leu Leu Leu Ala Ala Gly Val Gly Tle Tyr Leu Leu Pro Asn Arg l7760 atg agc acg aat cct aaa cct caa aga aag acc aaa cgt aac acc aac l8039 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn l775 Cgg cgg cgg cag gac gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt Arg Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly l790 gga gtt tac ttg ttg ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg l8135 acg aga aag act tcc gag cgg tgc caa act ccg ggt ggt ggt gcg gcg l8135 acg aga aag act tcc gag cgg tgc caa cct cga ggt aga cgt cag cct l81835 acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct l81835 acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct l81835 acc aca aag gct cgt cgc cga ggg ggc aga acc tgg gct cag cct l81835 atc ccc aag gct cgt cgc cga ggg aga cct gg gct cag ccc ggg l1e Pro Lys Ala Arg Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly l840 tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg gcg gga tgg l1e Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gly Trp l850 tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg gcg gga tgg l827 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp l855 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg gcc ca aca gac ccc Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro l870 1870	att tat cac agc gtg tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc 1 Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys 1740 1745 1750 1755	7943
Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn 1785 Cgg cgg ccg cag gac gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly 1790 1790 1790 1795 1800 1810 1815 Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 1805 1810 1815 acg aga aag act tcc agg cgg cgg caa cct cga ggt aga cgt cag cct 18183 acg aga aag act tcc agg cgg tcg caa cct cga ggt aga cgt cag cct 18183 atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg lle Pro 1820 1820 1820 1835 atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg lle Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 1840 1845 1855 ccc ttgg ccc ctc tat ggc aat gag ggc tgg ggg tgg gg gg atgg 1827 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 ctc ctg tct ccc ggt ggc tct cgg cct agc tgg ggc gg agg acc cc 282 Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1870 1875 1895 ggc ttc gcc gac ctc atg ggt ac atg ggt acc ct acc gt gc Arg Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 1890 1895 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc taatagtcga Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttcccgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgtac caactttaca 18541 catactttat atagctactc actcaagaga ttggactga cgagcaggc gggtgggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcacagac gggtgggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcacagac gggcaggac tggggtgggt 18721	Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg	7991
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly 1790 1795 1800 gga gtt tac ttg ttg ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 1805 1810 1810 1815 acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro 1820 1825 1830 1835 atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 1850 tac cct tgg ccc ctc tat ggc aat gag ggc tgg ggg tgg ggg ggg tgg Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 1860 1865 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1870 1880 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn	.8039
Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala 1805 1810 1815 1815 acg aga aag act tcc gag cgg tcg caa cct cga ggt aga cgt cag cct 18183 Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Gln Pro 1825 1830 1835 atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg 18231 Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 1845 1850 tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg ggg gga tgg 18279 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 1860 1865 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1870 1875 1880 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg	Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gln Ile Val Gly	.8087
Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro 1825 atc ccc aag gct cgt cgg ccc gag ggc agg acc tgg gct cag ccc ggg 18231 The Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly 1840 tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg ggg gga tgg 18279 Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 tcc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc 18327 Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1875 1875 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc 18375 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc taatagtcga 18421 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caacttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaaagagaa ttggatctga cgagcaggac tgggcggcgg 18781	Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala	8135
tac cct tgg ccc ctc tat ggc aat gag ggc tgc ggg tgg ggg gga tgg Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 1860 1865 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1875 1880 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 1890 1895 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc taatagtcga Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 1905 1910 ctttgttecc actgtacttt tagctcgtac aaaatacaat atactttca tttctcgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro	.8183
Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp 1855 1860 1865 ctc ctg tct ccc cgt ggc tct cgg cct agc tgg ggc ccc aca gac ccc Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1875 1880 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc Arg Arg Arg Arg Arg Asp Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 1890 1895 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc taatagtcga 18421 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 1905 1910 ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 18481 aacaacatgt tttcccatgt aatatecttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly	8231
Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro 1870 1875 1880 cgg cgt agg tcg cgc aat ttg ggt aag gtc atc gat acc ctt acg tgc 18375 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 1890 1895 ggc ttc gcc gac ctc atg ggg tac ata ccg ctc gtc taatagtcga 18421 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 1905 1910 ctttgttccc actgtacttt tagctcgtac aaaatacaat atactttca tttctccgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp	.8279
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys 1885 1890 1895 1895 1895 1895 18421 Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 1905 1910 ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro	.8327
Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val 1900 1905 1910 ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 18481 aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys	.8375
aacaacatgt tttcccatgt aatatccttt tctattttc gttccgttac caactttaca 18541 catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val	
catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 18601 tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	ctttgttccc actgtacttt tagctcgtac aaaatacaat atacttttca tttctccgta 1	.8481
tgcctgccat atttcaattt gttataaatt cctataattt atcctattag tagctaaaaa 18661 aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	aacaacatgt tttcccatgt aatatccttt tctatttttc gttccgttac caactttaca 1	.8541
aagatgaatg tgaatcgaat cctaagagaa ttggatctga tccacaggac gggtgtggtc 18721 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781	catactttat atagctattc acttctatac actaaaaaac taagacaatt ttaattttgc 1	8601
gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 18781		
	ccaaagcggt cggacagtgc tccgagaacg ggtgcgcata gaaattgcat caacgcatat 1	
agegetagea geacgecata gtgactggeg atgetgtegg aatggacgat atcccgcaag 18901		

 $\{ , _{2}, \cdot \}$

aggcccggca gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg 18961 aggatgacga tgagcgcatt gttagatttc atacacggtg cctgactgcg ttagcaattt 19021 aactgtgata aactaccgca ttaaagcttt ttctttccaa ttttttttt ttcgtcatta 19081 taaaaatcat tacgaccgag attcccgggt aataactgat ataattaaat tgaagctcta 19141 atttgtgagt ttagtataca tgcatttact tataatacag ttttttagtt ttgctggccg 19201 catcttctca aatatgcttc ccagcctgct tttctgtaac gttcaccctc taccttagca 19261 tcccttccct ttgcaaatag tcctcttcca acaataataa tgtcagatcc tgtagagacc 19321 acatcateca eggitetata etgitgacee aatgegiete eetigicate taaaceeaca 19381 ccgggtgtca taatcaacca atcgtaacct tcatctcttc cacccatgtc tctttgagca 19441 ataaagccga taacaaaatc tttgtcgctc ttcgcaatgt caacagtacc cttagtatat 19501 tctccagtag atagggagcc cttgcatgac aattctgcta acatcaaaag gcctctaggt 19561 tectttgtta ettettetge egeetgette aaacegetaa caatacetqg geecaceaca 19621 ccgtgtgcat tcgtaatgtc tgcccattct gctattctgt atacacccgc agagtactgc 19681 aatttgactg tattaccaat gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac 19741 ttggcggata atgcctttag cggcttaact gtgccctcca tggaaaaatc agtcaagata 19801 tccacatgtg tttttagtaa acaaattttg ggacctaatg cttcaactaa ctccagtaat 19861 tccttggtgg tacgaacatc caatgaagca cacaagtttg tttgcttttc gtgcatgata 19921 ttaaataget tggcagcaac aggactagga tgagtagcag cacgtteett atatgtaget 19981 ttcgacatga tttatcttcg tttcctgcag gtttttgttc tgtgcagttg ggttaagaat 20041 actgggcaat ttcatgtttc ttcaacacta catatgcgta tatataccaa tctaagtctg 20101 tgctccttcc ttcgttcttc cttctgttcg gagattaccg aatcaaaaaa atttcaagga 20161 aaccgaaatc aaaaaaaaga ataaaaaaaaa aatgatgaat tgaaaagctt atcgat

```
<210> 17
<211> 1911
<212> PRT
<213> Artificial Sequence
```

<220>
<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core140

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

- Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
 35 40 45
- Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 55 60
- Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80
- Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
 85 90 95
- Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110
- Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125
- Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140
- Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys 145 150 155 160
- Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala 165 170 175
- Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
 180 185 190
- Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe 195 200 205
- Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe 210 215 220
- Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp 225 230 235
- Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro 245 250 255
- Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe 260 265 270
- Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr 275 280 285
- Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn 290 295 300
- Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly 305 310 315 320
- Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr 325 330 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr 340 345 350

- Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp 355 360 365
- Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu 370 380
- Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro 385 390 395 400
- Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val 405 410 415
- Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala 420 425 430
- Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu 435 440 445
- Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu 450 455 460
- Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 465 470 475 480
- Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 495
- Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr 500 505 510
- Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 515 520 525
- Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn 530 535 540
- Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 545 550 555 560
- Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val 565 570 575
- Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
 580 585 590
- Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu 595 600 605
- Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 610 615 620
- Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val 625 630 635 640
- Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val 645 650 655

- Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala 660 665 670
- Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His 675 680 685
- Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Arg Val 690 695 700
- Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 705 710 715 720
- His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735
- Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750
- Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe 755 760 765
- Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile 770 780
- Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys 785 790 795 800
- Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815
- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845
- Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 860
- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875 880
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990

- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100
- Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120
- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr 1300 1305 1310

- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1335 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420
- Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440
- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630

- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1740
- Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala 745 1750 1755 1760
- Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro 1765 1770 1775
- Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp 1780 1785 1790
- Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu 1795 1800 1805
- Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser 1810 1815 1820
- Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg 825 1830 1835 1840
- Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu 1845 1850 1855
- Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg 1860 1865 1870
- Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg 1885
- Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu 1890 1895 1900
- Met Gly Tyr Ile Pro Leu Val 905 1910

<210> 18

```
<211> 20247
<212> DNA
<213> Artificial Sequence
<223> Description of Artificial Sequence:
     pd.delta.NS3NS5.pj.core150
<220>
<221> CDS
<222> (12679)..(18441)
<400> 18
ategatecta cecettgege taaagaagta tatgtgeeta etaacqettg tetttgtete 60
tgtcactaaa cactggatta ttactcccag atacttattt tggactaatt taaatgattt 120
cggatcaacg ttcttaatat cgctgaatct tccacaattg atgaaagtag ctaggaagag 180
gaattggtat aaagtttttg tttttgtaaa tctcgaagta tactcaaacg aatttagtat 240
tttctcagtg atctcccaga tgctttcacc ctcacttaga agtgctttaa gcatttttt 300
actgtggcta tttcccttat ctgcttcttc cgatgattcg aactgtaatt gcaaactact 360
tacaatatca gtgatatcag attgatgttt ttgtccatag taaggaataa ttgtaaattc 420
ccaagcagga atcaatttct ttaatgaggc ttccagaatt gttgcttttt gcgtcttgta 480
tttaaactgg agtgatttat tgacaatatc gaaactcagc gaattgctta tgatagtatt 540
atageteatg aatgtggete tettgattge tgtteegtta tgtgtaatea teeaacataa 600
ataggttagt tcagcagcac ataatgctat tttctcacct gaaggtcttt caaacctttc 660
cacaaactga cgaacaagca ccttaggtgg tgttttacat aatatatcaa attgtggcat 720
gettagegee gatettgtgt geaattgata tetagtttea actaetetat ttatettgta 780
tcttgcagta ttcaaacacg ctaactcgaa aaactaactt taattgtcct gtttgtctcg 840
cgttctttcg aaaaatgcac cggccgcgca ttatttgtac tgcgaaaata attggtactg 900
eggtatette attteatatt ttaaaaatge acetttgetg etttteetta atttttagae 960
ggcccgcagg ttcgttttgc ggtactatct tgtgataaaa agttgttttg acatgtqatc 1020
tgcacagatt ttataatgta ataagcaaga atacattatc aaacgaacaa tactggtaaa 1080
agaaaaccaa aatggacgac attgaaacag ccaagaatct gacggtaaaa gcacgtacag 1140
cttatagcgt ctgggatgta tgtcggctgt ttattgaaat gattgctcct gatgtagata 1200
ttgatataga gagtaaacgt aagtctgatg agctactctt tccaggatat gtcataaggc 1260
ccatggaatc tctcacaacc ggtaggccgt atggtcttga ttctagcgca gaagattcca 1320
gcgtatcttc tgactccagt gctgaggtaa ttttgcctgc tgcgaagatg gttaaggaaa 1380
```

ggtttgattc gattggaaat ggtatgctct cttcacaaga agcaagtcag gctgccatag 1440 atttgatgct acagaataac aagctgttag acaatagaaa gcaactatac aaatctattg 1500 ctataataat aggaagattg cccgagaaag acaagaagag agctaccgaa atgctcatga 1560 gaaaaatgga ttgtacacag ttattagtcc caccagctcc aacggaagaa gatgttatga 1620 agctcgtaag cgtcgttacc caattgctta ctttagttcc accagatcgt caagctgctt 1680 taataggtga tttattcatc ccggaatctc taaaggatat attcaatagt ttcaatgaac 1740 tggcggcaga gaatcgttta cagcaaaaaa agagtgagtt ggaaggaagg actgaagtga 1800 accatgetaa tacaaatgaa gaagtteeet eeaggegaac aagaagtaga gacacaaatg 1860 caagaggagc atataaatta caaaacacca tcactgaggg ccctaaagcg gttcccacga 1920 aaaaaaggag agtagcaacg agggtaaggg gcagaaaatc acgtaatact tctagggtat 1980 gatccaatat caaaggaaat gatagcattg aaggatgaga ctaatccaat tgaggagtgg 2040 cagcatatag aacagctaaa gggtagtgct gaaggaagca tacgataccc cgcatggaat 2100 gggataatat cacaggaggt actagactac ctttcatcct acataaatag acgcatataa 2160 gtacgcattt aagcataaac acgcactatg ccgttcttct catgtatata tatatacagg 2220 caacacgcag atataggtgc gacgtgaaca gtgagctgta tgtgcgcagc tcgcgttgca 2280 ttttcggaag cgctcgtttt cggaaacgct ttgaagttcc tattccgaag ttcctattct 2340 ctagaaagta taggaacttc agagcgcttt tgaaaaccaa aagcgctctq aagacqcact 2400 ttcaaaaaac caaaaacgca ccggactgta acgagctact aaaatattgc gaataccgct 2460 tccacaaaca ttgctcaaaa gtatctcttt gctatatatc tctgtgctat atccctatat 2520 aacctaccca tccacctttc gctccttgaa cttgcatcta aactcgacct ctacatcaac 2580 aggetteeaa tgetetteaa attttaetgt caagtagace cataeggetg taatatgetg 2640 ctcttcataa tgtaagctta tctttatcga atcgtgtgaa aaactactac cgcgataaac 2700 ctttacggtt ccctgagatt gaattagttc ctttagtata tgatacaaga cacttttgaa 2760 ctttgtacga cgaattttga ggttcgccat cctctggcta tttccaatta tcctgtcggc 2820 tattatetee geeteagttt gatetteege tteagaetge cattttteae ataatgaate 2880 tatttcaccc cacaatcctt catccgcctc cgcatcttgt tccgttaaac tattgacttc 2940 atgttgtaca ttgtttagtt cacgagaagg gtcctcttca ggeggtagct cctgatctcc 3000 tatatgacct ttatcctgtt ctctttccac aaacttagaa atgtattcat gaattatgga 3060 gcacctaata acattettea aggeggagaa gtttgggeea gatgeecaat atgettgaea 3120 tgaaaacgtg agaatgaatt tagtattatt gtgatattct gaggcaattt tattataatc 3180

tcgaagataa gagaagaatg cagtgacctt tgtattgaca aatggagatt ccatgtatct 3240 aaaaaatacq cctttaggcc ttctgatacc ctttcccctg cggtttagcg tgccttttac 3300 attaatatet aaaccetete egatggtgge etttaaetga etaataaatg caaccgatat 3360 aggatcaggc caatccagtt ctttttcaat taccggtgtg tcgtctgtat tcagtacatg 3480 tccaacaaat gcaaatgcta acgttttgta tttcttataa ttgtcaggaa ctggaaaagt 3540 cccccttqtc gtctcgatta cacacctact ttcatcgtac accataggtt ggaagtgctg 3600 cataatacat tgcttaatac aagcaagcag tctctcgcca ttcatatttc agttattttc 3660 cattacagct gatgtcattg tatatcagcg ctgtaaaaat ctatctgtta cagaaggttt 3720 tegeggtttt tataaacaaa actttegtta egaaategag caateacece agetgegtat 3780 ttggaaattc gggaaaaagt agagcaacgc gagttgcatt ttttacacca taatgcatga 3840 ttaacttcga gaagggatta aggctaattt cactagtatg tttcaaaaaac ctcaatctgt 3900 ccattgaatg ccttataaaa cagctataga ttgcatagaa gagttagcta ctcaatgctt 3960 tttgtcaaag cttactgatg atgatgtgtc tactttcagg cgggtctgta gtaaggagaa 4020 tgacattata aagctggcac ttagaattcc acggactata gactatacta gtatactccg 4080 totactgtac gatacacttc cgctcaggtc cttgtccttt aacgaggcct taccactctt 4140 ttgttactct attgatccag ctcagcaaag gcagtgtgat ctaagattct atcttcgcga 4200 tgtagtaaaa ctagctagac cgagaaagag actagaaatg caaaaggcac ttctacaatg 4260 gctgccatca ttattatccg atgtgacgct gcattttttt ttttttttt tttttttt 4320 tttttttttt tttttttt ttttttggta caaatatcat aaaaaaagag aatcttttta 4380 agcaaggatt ttcttaactt cttcggcgac agcatcaccg acttcggtgg tactgttgga 4440 accacctaaa tcaccagttc tgatacctgc atccaaaacc tttttaactg catcttcaat 4500 ggctttacct tcttcaggca agttcaatga caatttcaac atcattgcag cagacaagat 4560 agtggcgata gggttgacct tattctttgg caaatctgga gcggaaccat ggcatggttc 4620 gtacaaacca aatgcggtgt tcttgtctgg caaagaggcc aaggacgcag atggcaacaa 4680 acccaaggag cctgggataa cggaggcttc atcggagatg atatcaccaa acatgttgct 4740 ggtgattata ataccattta ggtgggttgg gttcttaact aggatcatgg cggcagaatc 4800 aatcaattga tgttgaactt tcaatgtagg gaattcgttc ttgatggttt cctccacagt 4860 ttttctccat aatcttgaag aggccaaaac attagcttta tccaaggacc aaataggcaa 4920 tggtggctca tgttgtaggg ccatgaaagc ggccattctt gtgattcttt gcacttctgg 4980

aacggtgtat tgttcactat cccaagcgac accatcacca tcgtcttcct ttctcttacc 5040 aaagtaaata cctcccacta attctctaac aacaacgaag tcagtacctt tagcaaattg 5100 tggcttgatt ggagataagt ctaaaagaga gtcggatgca aagttacatg gtcttaagtt 5160 ggcgtacaat tgaagttett tacggatttt tagtaaacct tgttcaggte taacactace 5220 ggtaccccat traggaccac ccacagcacc taacaaaacg gcatcagcct tcttggaggc 5280 ttccagcgcc tcatctggaa gtggaacacc tgtagcatcg atagcagcac caccaattaa 5340 atgattttcg aaatcgaact tgacattgga acgaacatca gaaatagctt taagaacctt 5400 aatggcttcg gctgtgattt cttgaccaac gtggtcacct ggcaaaacga cgatcttctt 5460 aaaaaaaaa atgcagcttc tcaatgatat tcgaatacgc tttgaggaga tacagcctaa 5580 tatccgacaa actgttttac agatttacga tcgtacttgt tacccatcat tgaattttga 5640 acatccgaac ctgggagttt tccctgaaac agatagtata tttgaacctg tataataata 5700 tatagtctag cgctttacgg aagacaatgt atgtatttcg gttcctggag aaactattgc 5760 atctattgca taggtaatct tgcacgtcgc atccccggtt cattttctgc gtttccatct 5820 tgcacttcaa tagcatatct ttgttaacga agcatctgtg cttcattttg tagaacaaaa 5880 atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 5940 gaaatgcaac gcgaaagcgc tattttacca acgaagaatc tgtgcttcat ttttgtaaaa 6000 caaaaaatgca acgcgagagc gctaattttt caaacaaaga atctgagctg catttttaca 6060 gaacagaaat gcaacgcgag agcgctattt taccaacaaa gaatctatac ttcttttttq 6120 ttctacaaaa atgcatcccg agagcgctat ttttctaaca aagcatctta gattactttt 6180 tttctccttt gtgcgctcta taatgcagtc tcttgataac tttttgcact gtaggtccgt 6240 taaggttaga agaaggctac tttggtgtct attttctctt ccataaaaaa agcctgactc 6300 cacttcccgc gtttactgat tactagcgaa gctgcgggtg catttttca agataaaggc 6360 atccccgatt atattctata ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata 6420 gcgttgatga ttcttcattg gtcagaaaat tatgaacggt ttcttctatt ttgtctctat 6480 atactacgta taggaaatgt ttacattttc gtattgtttt cgattcactc tatgaatagt 6540 tcttactaca attttttgt ctaaagagta atactagaga taaacataaa aaatgtagag 6600 gtcgagttta gatgcaagtt caaggagcga aaggtggatg ggtaggttat atagggatat 6660 agcacagaga tatatagcaa agagatactt ttgagcaatg tttgtggaag cggtattcgc 6720 aatattttag tagctcgtta cagtccggtg cgtttttggt tttttgaaag tgcgtcttca 6780

gagegetttt ggtttteaaa agegetetga agtteetata etttetagag aataggaaet 6840 teggaatagg aactteaaag egttteegaa aacgageget teegaaaatg caacgegage 6900 tgcgcacata cageteactg ttcacgtege acetatatet gegtgttgce tgtatatata 6960 tatacatgag aagaacggca tagtgcgtgt ttatgcttaa atgcgtactt atatgcgtct 7020 atttatgtag gatgaaaggt agtctagtac ctcctgtgat attatcccat tccatgcggg 7080 gtategtatg etteetteag cactaceett tagetgttet atatgetgee acteeteaat 7140 tggattagtc tcatccttca atgctatcat ttcctttgat attggatcat atgcatagta 7200 ccgagaaact agtgcgaagt agtgatcagg tattgctgtt atctgatgag tatacgttgt 7260 cetggccacg geagaageae gettateget ceaattteee acaacattag teaacteegt 7320 taggcccttc attgaaagaa atgaggtcat caaatgtctt ccaatgtgag attttgggcc 7380 attttttata gcaaagattg aataaggcgc atttttcttc aaagctttat tgtacgatct 7440 gactaagtta tettttaata attggtatte etgtttattg ettgaagaat tgeeggteet 7500 atttactcgt tttaggactg gttcagaatt cctcaaaaat tcatccaaat atacaagtgg 7560 atcgatgata agctgtcaaa catgagaatt cttgaagacg aaagggcctc gtgatacgcc 7620 tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 7680 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 7740 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 7800 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 7860 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 7920 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 7980 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 8040 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 8100 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 8160 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 8220 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 8280 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 8340 cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 8400 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 8460 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 8520 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 8580

cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 8640 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 8700 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 8760 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 8820 gatettettg agateetttt tttetgegeg taatetgetg ettgeaaaca aaaaaaccae 8880 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa 8940 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 9000 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 9060 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 9120 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 9180 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 9240 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 9300 cgagggagct tccaggggga aacgcctggt atctttatag tcctgtcggg tttcgccacc 9360 tetgaettga gegtegattt ttgtgatget egteaggggg geggageeta tggaaaaaeg 9420 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 9480 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 9540 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 9600 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatggtgca 9660 ctctcagtac aatctgctct gatgccgcat agttaagcca gtatacactc cgctatcgct 9720 acgtgactgg gtcatggctg cgccccgaca cccgccaaca cccgctgacg cgccctgacg 9780 ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 9840 gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagctgcggt aaagctcatc 9900 agcgtggtcg tgaagcgatt cacagatgtc tgcctgttca tccgcgtcca gctcgttgag 9960 tttctccaga agcgttaatg tctggcttct gataaagcgg gccatgttaa gggcggtttt 10020 ttcctgtttg gtcactgatg cctccgtgta agggggattt ctgttcatgg gggtaatgat 10080 accgatgaaa cgagagagga tgctcacgat acgggttact gatgatgaac atgcccggtt 10140 actggaacgt tgtgagggta aacaactggc ggtatggatg cggcgggacc agagaaaaat 10200 cactcagggt caatgccagc gcttcgttaa tacagatgta ggtgttccac agggtagcca 10260 gcagcatcct gcgatgcaga tccggaacat aatggtgcag ggcgctgact tccgcgtttc 10320 cagactttac gaaacacgga aaccgaagac cattcatgtt gttgctcagg tcgcagacgt 10380

tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt gattcattct gctaaccagt 10440 aaggcaaccc cgccagccta gccgggtcct caacgacagg agcacgatca tgcgcacccg 10500 tggccaggac ccaacgetgc ccgagatgcg ccgcgtgcgg ctgctggaga tggcggacgc 10560 gatggatatg ttctgccaag ggttggtttg cgcattcaca gttctccgca agaattgatt 10620 ggctccaatt cttggagtgg tgaatccgtt agcgaggtgc cgccggcttc cattcaggtc 10680 gaggtggccc ggctccatgc accgcgacgc aacgcgggga ggcagacaag gtatagggcg 10740 gegectacaa tecatgecaa ecegttecat gtgetegeeg aggeggeata aategeegtg 10800 acgatcagcg gtccaatgat cgaagttagg ctggtaagag ccgcgagcga tccttgaagc 10860 tgtccctgat ggtcgtcatc tacctgcctg gacagcatgg cctgcaacgc gggcatcccg 10920 atgccgccgg aagcgagaag aatcataatg gggaaggcca tccagcctcg cgtcgcgaac 10980 gccagcaaga cgtagcccag cgcgtcggcc gccatgccgg cgataatggc ctgcttctcg 11040 ccgaaacgtt tggtggcggg accagtgacg aaggcttgag cgagggcgtg caagattccg 11100 aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc gaaagcggtc ctcgccgaaa 11160 atgacccaga gcgctgccgg cacctgtcct acgagttgca tgataaagaa gacagtcata 11220 agtgcggcga cgatagtcat gccccgcgcc caccggaagg agctgactgg gttgaaggct 11280 ctcaagggca teggtegagg atcettcaat atgegeacat acgetgttat gttcaaggte 11340 ccttcgttta agaacgaaag cggtcttcct tttgagggat gtttcaagtt gttcaaatct 11400 atcaaatttg caaatcccca gtctgtatct agagcgttga atcggtgatg cgatttgtta 11460 attaaattga tggtgtcacc attaccaggt ctagatatac caatggcaaa ctgagcacaa 11520 caataccagt ceggateaac tggcaccate tetecegtag teteatetaa tttttettee 11580 ggatgaggtt ccagatatac cgcaacacct ttattatggt ttccctgagg gaataataqa 11640 atgtcccatt cgaaatcacc aattctaaac ctgggcgaat tgtatttcgg gtttgttaac 11700 tegttecagt caggaatgtt ccacgtgaag ctatetteca gcaaagtete caettettea 11760 tcaaattgtg gagaatactc ccaatgctct tatctatggg acttccggga aacacagtac 11820 cgatacttcc caattcgtct tcagagctca ttgtttgttt gaagagacta atcaaagaat 11880 cgttttctca aaaaaattaa tatcttaact gatagtttga tcaaaggggc aaaacgtagg 11940 ggcaaacaaa cggaaaaatc gtttctcaaa ttttctgatg ccaagaactc taaccagtct 12000 tatctaaaaa ttgccttatg atccgtctct ccggttacag cctgtgtaac tgattaatcc 12060 tgcctttcta atcaccattc taatgtttta attaagggat tttgtcttca ttaacggctt 12120 tegetcataa aaatgttatg aegttttgee egeaggeggg aaaccateea etteaegaga 12180

ctgatctcct ctgccg	gaac accgggcat	c tccaacttat	aagttggaga aataa	ngagaa 12240
tttcagattg agagaa	tgaa aaaaaaaa	c ccttagttca	taggtccatt ctctt	agcgc 12300
aactacagag aacagg	ggca caaacaggc	a aaaaacgggc	acaacctcaa tggag	stgatg 12360
caacctgcct ggagta	aatg atgacacaa	g gcaattgacc	cacgcatgta tctat	ctcat 12420
tttcttacac cttcta	ttac cttctgctc	t ctctgatttg	gaaaaagctg aaaaa	aaaagg 12480
ttgaaaccag ttccct	gaaa ttattcccc	t acttgactaa	taagtatata aagad	ggtag 12540
gtattgattg taattc	tgta aatctattt	c ttaaacttct	taaattctac tttta	atagtt 12600
agtctttttt ttagtt	ttaa aacaccaag	a acttagtttc	gaataaacac acata	naacaa 12660
acaagcttac aaaaca			cag ggc tat aag Gln Gly Tyr Lys 10	
cta gta ctc aac c Leu Val Leu Asn P 15	cc tct gtt gct ro Ser Val Ala	gca aca ctg Ala Thr Leu 20	ggc ttt ggt gct Gly Phe Gly Ala 25	tac 12759 Tyr
atg tcc aag gct c Met Ser Lys Ala H 30	at ggg atc gat is Gly Ile Asp 35	Pro Asn Ile	agg acc ggg gtg Arg Thr Gly Val 40	aga 12807 Arg
aca att acc act g Thr Ile Thr Thr G 45	gc agc ccc atc ly Ser Pro Ile 50	acg tac tcc Thr Tyr Ser	acc tac ggc aag Thr Tyr Gly Lys 55	ttc 12855 Phe
ctt gcc gac ggc g Leu Ala Asp Gly G 60				
gac gag tgc cac t Asp Glu Cys His S	cc acg gat gcc er Thr Asp Ala 80	aca tcc atc Thr Ser Ile 85	ttg ggc att ggc Leu Gly Ile Gly 90	act 12951 Thr
gtc ctt gac caa g Val Leu Asp Gln A 95	ca gag act gcg la Glu Thr Ala	ggg gcg aga Gly Ala Arg 100	ctg gtt gtg ctc Leu Val Val Leu 105	gcc 12999 Ala
acc gcc acc cct c Thr Ala Thr Pro P 110	cg ggc tcc gtc ro Gly Ser Val 115	Thr Val Pro	cat ccc aac atc His Pro Asn Ile 120	gag 13047 Glu
gag gtt gct ctg t Glu Val Ala Leu S 125	cc acc acc gga er Thr Thr Gly 130	gag atc cct Glu Ile Pro	ttt tac ggc aag Phe Tyr Gly Lys 135	gct 13095 Ala
atc ccc ctc gaa g Ile Pro Leu Glu V 140	ta atc aag ggg al Ile Lys Gly 145	ggg aga cat Gly Arg His 150	ctc atc ttc tgt Leu Ile Phe Cys	cat 13143 His 155
tca aag aag t Ser Lys Lys Lys C 1	gc gac gaa ctc ys Asp Glu Leu 60	gcc gca aag Ala Ala Lys 165	ctg gtc gca ttg Leu Val Ala Leu 170	ggc 13191 Gly

		gcc Ala														13239
		ggc Gly 190														13287
		ggc Gly														13335
		gtc Val														13383
acg Thr	ctc Leu	ccc Pro	caa Gln	gat Asp 240	gct Ala	gtc Val	tcc Ser	cgc Arg	act Thr 245	caa Gln	cgt Arg	cgg Arg	ggc Gly	agg Arg 250	act Thr	13431
ggc Gly	agg Arg	gly ggg	aag Lys 255	cca Pro	ggc	atc Ile	tac Tyr	aga Arg 260	ttt Phe	gtg Val	gca Ala	ccg Pro	ggg Gly 265	gag Glu	cgc Arg	13479
		ggc Gly 270														13527
ggc Gly	tgt Cys 285	gct Ala	tgg Trp	tat Tyr	gag Glu	ctc Leu 290	acg Thr	ccc Pro	gcc Ala	gag Glu	act Thr 295	aca Thr	gtt Val	agg Arg	cta Leu	13575
		tac Tyr														13623
		tgg Trp														13671
ttt Phe	cta Leu	tcc Ser	cag Gln 335	aca Thr	aag Lys	cag Gln	agt Ser	ggg Gly 340	Glu	Asn	ctt Leu	Pro	Tyr	Leu	gta Val	13719
gcg Ala	tac Tyr	caa Gln 350	gcc Ala	acc Thr	gtg Val	tgc Cys	gct Ala 355	agg Arg	gct Ala	caa Gln	gcc Ala	cct Pro 360	ccc Pro	cca Pro	tcg Ser	13767
tgg Trp	gac Asp 365	cag Gln	atg Met	tgg Trp	aag Lys	tgt Cys 370	ttg Leu	att Ile	cgc Arg	ctc Leu	aag Lys 375	ccc Pro	acc Thr	ctc Leu	cat His	13815
380 380	cca Pro	aca Thr	ccc Pro	ctg Leu	cta Leu 385	tac Tyr	aga Arg	ctg Leu	ggc Gly	gct Ala 390	gtt Val	cag Gln	aat Asn	gaa Glu	atc Ile 395	13863
acc Thr	ctg Leu	acg Thr	cac His	cca Pro 400	gtc Val	acc Thr	aaa Lys	tac Tyr	atc Ile 405	atg Met	aca Thr	tgc Cys	atg Met	tcg Ser 410	gcc Ala	13911

		gag Glu														13959
		ttg Leu 430														14007
		gtc Val														14055
-		tac Tyr	-	-				_	-		_		_			14103
_		atc Ile				_	_		_		_		_	_	_	14151
		ggc Gly														14199
		gtc Val 510														14247
		tgg Trp														14295
		cct Pro														14343
		acc Thr														14391
		Gly 999		Val	Ala		Gln	Leu	Ala					Ala		14439
gcc Ala	ttt Phe	gtg Val 590	ggc Gly	gct Ala	ggc Gly	tta Leu	gct Ala 595	ggc Gly	gcc Ala	gcc Ala	atc Ile	ggc Gly 600	agt Ser	gtt Val	gga Gly	14487
ctg Leu	999 605	aag Lys	gtc Val	ctc Leu	ata Ile	gac Asp 610	atc Ile	ctt Leu	gca Ala	ggg ggg	tat Tyr 615	ggc Gly	gcg Ala	ggc Gly	gtg Val	14535
gcg Ala 620	gga Gly	gct Ala	ctt Leu	gtg Val	gca Ala 625	ttc Phe	aag Lys	atc Ile	atg Met	agc Ser 630	ggt Gly	gag Glu	gtc Val	ccc Pro	tcc Ser 635	14583
acg Thr	gag Glu	gac Asp	ctg Leu	gtc Val 640	aat Asn	cta Leu	ctg Leu	ccc Pro	gcc Ala 645	atc Ile	ctc Leu	tcg Ser	ccc Pro	gga Gly 650	gcc Ala	14631

			ggc Gly 655													14679
			G1Å 888													14727
			aac Asn													14775
_	-		cgc Arg	-		_				_			_		_	14823
	_		cga Arg	_		_			-	_		_				14871
_			tcc Ser 735							_			_			14919
_	_		ttt Phe	_					_	_		_		_	_	14967
			ccc Pro													15015
_		-	ggc Gly		_			_	_		_		-			15063
			gtc Val					_			_					15111
_		Asn	atg Met 815	Trp	Ser	Gly	Thr	Phe	Pro	Ile	Asn	Ăla	Tyr	Thr	_	15159
			acc Thr													15207
			gca Ala													15255
			acg Thr													15303
gtc Val			ccc Pro													15351

agg ttt gcg ccc ccc tgc aag ccc ttg ctg cgg gag gag gta tca ttc Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe 895 900 905	
aga gta gga ctc cac gaa tac ccg gta ggg tcg caa tta cct tgc gag Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu 910 915 920	
CCC gaa CCg gac gtg gcc gtg ttg acg tcc atg ctc act gat CCC tcc Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser 925 930 935	
cat ata aca gca gag gcg gcc ggg cga agg ttg gcg agg gga tca ccc His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro 940 945 950)
ccc tct gtg gcc agc tcc tcg gct agc cag cta tcc gct cca tct ctc Pro Ser Val Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu 960 965 970	
aag gca act tgc acc gct aac cat gac tcc cct gat gct gag ctc ata Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile 975 980 985	
gag gcc aac ctc cta tgg agg cag gag atg ggc ggc aac atc acc agg Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg 990 995 1000	
gtt gag tca gaa aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu 1005 1010 1015	
gtg gcg gag gag gac gag cgg gag atc tcc gta ccc gca gaa atc ctc Val Ala Glu Glu Asp Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu 1020 1025 1030	1
cgg aag tot cgg aga tto goo cag goo otg coo gtt tgg gog cgg coo Arg Lys Ser Arg Arg Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro 1040 1045 1050	
gac tat aac ccc ccg cta gtg gag acg tgg aaa aag ccc gac tac gas Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu 1055 1060 1065	
cca cct gtg gtc cat ggc tgc ccg ctt cca cct cca aag tcc cct cct Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro 1070 1080	
gtg cct ccg cct cgg aag aag cgg acg gtg gtc ctc act gaa tca acc Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr 1085 1090 1095	
cta tct act gcc ttg gcc gag ctc gcc acc aga agc ttt ggc agc tcc Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser 1100 1105 1110 1115	5
tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro 1120 1125 1130	

gcc cct tct ggc tgc ccc ccc gac tcc gac gct gag tcc tat tcc tcc Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser 1135 1140 1145	16119
atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctt agc gac ggg Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly 1150 1155 1160	16167
tca tgg tca acg gtc agt agt gag gcc aac gcg gag gat gtc gtg tgc Ser Trp Ser Thr Val Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys 1165 1170 1175	16215
tgc tca atg tct tac tct tgg aca ggc gca ctc gtc acc ccg tgc gcc Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala 1180 1185 1190 1195	16263
gcg gaa gaa cag aaa ctg ccc atc aat gca cta agc aac tcg ttg cta Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1200 1205 1210	16311
cgt cac cac aat ttg gtg tat tcc acc acc tca cgc agt gct tgc caa Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln 1215 1220 1225	16359
agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His 1230 1235 1240	16407
tac cag gac gta ctc aag gag gtt aaa gca gcg gcg tca aaa gtg aag Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys 1245 1250 1255	16455
gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cac Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His 1260 1265 1270 1275	16503
tca gcc aaa tcc aag ttt ggt tat ggg gca aaa gac gtc cgt tgc cat Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His 1280 1285 1290	16551
gcc aga aag gcc gta acc cac atc aac tcc gtg tgg aaa gac ctt ctg Ala Arg Lys Ala Val Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu 1295 1300 1305	16599
gaa gac aat gta aca cca ata gac act acc atc atg gct aag aac gag Glu Asp Asn Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1310 1315 1320	16647
gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1325 1330 1335	16695
atc gtg ttc ccc gat ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1340 1345 1350 1355	16743
tac gac gtg gtt aca aag ctc ccc ttg gcc gtg atg gga agc tcc tac Tyr Asp Val Val Thr Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr 1360 1365 1370	16791

!* 1

4.3

gga ttc caa tac Gly Phe Gln Tyr 1375			Glu Phe Leu		16839
tgg aag tcc aag Trp Lys Ser Lys 1390	Lys Thr Pro				16887
ttt gac tcc aca Phe Asp Ser Thr 1405					16935
tac caa tgt tgt Tyr Gln Cys Cys 1420		Pro Gln Ala			16983
ctc acc gag agg Leu Thr Glu Arg			Leu Thr Asn		17031
gag aac tgc ggc Glu Asn Cys Gly 1455			Ser Gly Val	_	17079
agc tgt ggt aac Ser Cys Gly Asn 1470	Thr Leu Thr				17127
cga gcc gca ggg Arg Ala Ala Gly 1485		_			17175
tta gtc gtt atc Leu Val Val Ile 1500					17223
ctg aga gcc ttc Leu Arg Ala Phe			Tyr Ser Ala		17271
gac ccc cca caa Asp Pro Pro Gln 1535			Leu Ile Thr		17319
tcc aac gtg tca Ser Asn Val Ser 1550	Val Ala His				17367
ctc acc cgt gac Leu Thr Arg Asp 1565					17415
gca aga cac act Ala Arg His Thr 1580					17463
gcc ccc aca ctg Ala Pro Thr Leu		-	Met Thr His	_	17511

gtc ctt ata gcc ag Val Leu Ile Ala Ar 1615	g Asp Gln Leu			
tac ggg gcc tgc ta Tyr Gly Ala Cys Ty 1630		Pro Leu Asp Leu		
caa aga ctc cat gg Gln Arg Leu His Gl 1645				
ggt gaa atc aat ag Gly Glu Ile Asn Ar 1660				1
ccc ttg cga gct tg Pro Leu Arg Ala Ti 168	p Arg His Arg			
ctg gcc aga gga gg Leu Ala Arg Gly Gl 1695	y Arg Ala Ala			
tgg gca gta aga ac Trp Ala Val Arg Th 1710		Leu Thr Pro Ile		
cag ctg gac ttg to Gln Leu Asp Leu Se 1725				
att tat cac agc gu Ile Tyr His Ser Va 1740				3
cta ctc ctg ctt go Leu Leu Leu Leu A 170	la Ala Gly Val			
atg agc acg aat co Met Ser Thr Asn P 1775	co Lys Pro Gln		Arg Asn Thr Asi	
cgg cgg ccg cag ga Arg Arg Pro Gln Aa 1790		Pro Gly Gly Gly		
gga gtt tac ttg to Gly Val Tyr Leu Lo 1805				
acg aga aag act to Thr Arg Lys Thr So 1820)
atc ccc aag gct co Ile Pro Lys Ala A	rg Arg Pro Glu			

				ctc Leu		-				_					tgg Trp	18279
		1	1855				1	1860					1865			
															ccc	18327
Leu		Ser 1870	Pro	Arg	GIY		Arg 1875	Pro	ser	Trp	-	Pro 1880	Thr	Asp	Pro	
						•					•					
															tgc	18375
	Arg 1885	Arg	Ser	Arg		ьеи 1890	GIY	гàз	vaı		Asp 1895	Thr	Leu	Thr	Cys	
_					_					-						
		_	_		_				_		_		_		ctt	18423
1900		Ala	Asp	Leu	Met 1905	GIY	Tyr	11e		Leu 1910	Val	GIA	Ala		Leu 1915	
1500	•			•					•					•	1713	
_	_	_	_	agg	_	taat	agto	ga o	ettte	gttc	cc a	ctgta	actt	t		18471
GIÀ	GIĄ	Ата		Arg 1920	Ala											
			-													
+																10531

tagetegtae aaaatacaat ataettttea ttteteegta aacaacatgt ttteecatgt 18531 aatatccttt tctatttttc gttccgttac caactttaca catactttat atagctattc 18591 acttctatac actaaaaaac taagacaatt ttaattttgc tgcctgccat atttcaattt 18651 gttataaatt cctataattt atcctattag tagctaaaaa aagatgaatg tgaatcgaat 18711 cctaagagaa ttggatctga tccacaggac gggtgtggtc gccatgatcg cgtagtcgat 18771 agtggctcca agtagcgaag cgagcaggac tgggcggcgg ccaaagcggt cggacagtgc 18831 tccgagaacg ggtgcgcata gaaattgcat caacgcatat agcgctagca gcacgccata 18891 gtgactggcg atgctgtcgg aatggacgat atcccgcaag aggcccggca gtaccggcat 18951 aaccaagcct atgcctacag catccagggt gacggtgccg aggatgacga tgagcgcatt 19011 gttagatttc atacacggtg cctgactgcg ttagcaattt aactgtgata aactaccgca 19071 ttaaagcttt ttctttccaa ttttttttt ttcgtcatta taaaaatcat tacgaccgag 19131 attcccgggt aataactgat ataattaaat tgaagctcta atttgtgagt ttagtataca 19191 tgcatttact tataatacag ttttttagtt ttgctggccg catcttctca aatatgcttc 19251 tectetteca acaataataa tgteagatee tgtagagaee acateateea eggttetata 19371 ctgttgaccc aatgcgtctc ccttgtcatc taaacccaca ccgggtgtca taatcaacca 19431 atcgtaacct tcatctcttc cacccatgtc tctttgagca ataaagccga taacaaaatc 19491 tttgtcgctc ttcgcaatgt caacagtacc cttagtatat tctccagtag atagggagcc 19551 cttgcatgac aattctgcta acatcaaaag gcctctaggt tcctttgtta cttcttctgc 19611 egectgette aaacegetaa caatacetgg geceaceaea eegtgtgeat tegtaatgte 19671

tgcccattct gctattctgt atacacccgc agagtactgc aatttgactg tattaccaat 19731 gtcagcaaat tttctgtctt cgaagagtaa aaaattgtac ttggcggata atgcctttag 19791 cggcttaact gtgccctcca tggaaaaatc agtcaagata tccacatgtg tttttagtaa 19851 acaaattttg ggacctaatg cttcaactaa ctccagtaat tccttggtgg tacgaacatc 19911 caatgaagca cacaagtttg tttgctttc gtgcatgata ttaaatagct tggcagcaac 19971 aggactagga tgagtagcag cacgttcctt atatgtagct ttcgacatga tttatcttcg 20031 tttcctgcag gtttttgttc tgtgcagttg ggttaagaat actgggcaat ttcatgttc 20091 ttcaacaca catatgcgta tatataccaa tctaagtctg tgctccttcc ttcgttctc 20151 cttctgttcg gagattaccg aatcaaaaaa atttcaagga aaccgaaatc aaaaaaaaga 20211 ataaaaaaaa aatgatgaat tgaaaagctt atcgat

<210> 19

<211> 1921

<212> PRT

<213> Artificial Sequence

<220:

<223> Description of Artificial Sequence:
 pd.delta.NS3NS5.pj.core150

<400> 19

Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro 1 5 10 15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His 20 25 30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly 50 . 55 60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser 65 70 75 80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala 85 90 95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro 100 105 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser 115 120 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val 130 135 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys

Asp	Glu	Leu	Ala	Ala 165	Lys	Leu	Val	Ala	Leu 170	Gly	Ile	Asn	Ala	Val 175	Ala
Tyr	Tyr	Arg	Gly 180	Leu	Asp	Val	Ser	Val 185	Ile	Pro	Thr	Ser	Gly 190	Asp	Val
Val	Val	Val 195	Ala	Thr	Asp	Ala	Leu 200	Met	Thr	Gly	Tyr	Thr 205	Gly	Asp	Phe
qaA	Ser 210	Val	Ile	Asp	Сув	Asn 215	Thr	Cys	Val	Thr	Gln 220	Thr	Val	Asp	Phe
Ser 225	Leu	Asp	Pro	Thr	Phe 230	Thr	Ile	Glu	Thr	Ile 235	Thr	Leu	Pro	Gln	Asp 240
Ala	Val	Ser	Arg	Thr 245	Gln	Arg	Arg	Gly	Arg 250	Thr	Gly	Arg	Gly	Lys 255	Pro
Gly	Ile	Tyr	Arg 260	Phe	Val	Ala	Pro	Gly 265	Glu	Arg	Pro	Ser	Gly 270	Met	Phe
Asp	Ser	Ser 275	Val	Leu	Сув	Glu	Cys 280	Tyr	Asp	Ala	Gly	Сув 285	Ala	Trp	Tyr
Glu	Leu 290	Thr	Pro	Ala	Glu	Thr 295	Thr	Val	Arg	Leu	Arg 300	Ala	Tyr	Met	Asn
Thr 305	Pro	Gly	Leu	Pro	Val 310	Сув	Gln	Asp	His	Leu 315	Glu	Phe	Trp	Glu	Gly 320
Val	Phe	Thr	Gly	Leu 325	Thr	His	Ile	Asp	Ala 330	His	Phe	Leu	Ser	Gln 335	Thr
Lys	Gln	Ser	Gly 340	Glu	Asn	Leu	Pro	Tyr 345	Leu	Val	Ala	Tyr	Gln 350	Ala	Thr
Val	Сув	Ala 355	Arg	Ala	Gln	Ala	Pro 360	Pro	Pro	Ser	Trp	Asp 365	Gln	Met	Trp
Lys	Суs 370	Leu	Ile	Arg	Leu	Lys 375	Pro	Thr	Leu	His	Gly 380	Pro	Thr	Pro	Leu
Leu 385	Tyr	Arg	Leu	Gly	Ala 390	Val	Gln	Asn	Glu	Ile 395	Thr	Leu	Thr	His	Pro 400
Val	Thr	Lys	Tyr	Ile 405	Met	Thr	Сув	Met	Ser 410	Ala	Asp	Leu	Glu	Val 415	Val
Thr	Ser	Thr	Trp 420	Val	Leu	Val	Gly	Gly 425	Val	Leu	Ala	Ala	Leu 430	Ala	Ala
Tyr	Сув	Leu 435	Ser	Thr	Gly	Сув	Val 440	Val	Ile	Val	Gly	Arg 445	Val	Val	Leu
Ser	Gly 450	ГÀЗ	Pro	Ala	Ile	Ile 455	Pro	Asp	Arg	Glu	Val 460	Leu	Tyr	Arg	Glu

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln 465 470 475 480

- Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu 485 490 495
- Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr 500 505 510
- Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe 515 520 525
- Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn 530 540
- Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro 545 550 555 560
- Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
 565 570 575
- Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala 580 585 590
- Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
 595 600 605
- Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val 610 615 620
- Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val 625 630 635 640
- Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
 645 650 655
- Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala 660 665 670
- Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His 675 680 685
- Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Arg Val 690 695 700
- Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu 705 710 715 720
- His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
 725 730 735
- Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
 740 745 750
- Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe 755 760 765
- Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile 770 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys 785 790 795 800

- Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp 805 810 815
- Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro 820 825 830
- Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu 835 840 845
- Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly 850 860
- Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu 865 870 875 880
- Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro 885 890 895
- Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His 900 905 910
- Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val 915 920 925
- Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu 930 935 940
- Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser 945 950 955 960
- Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr 965 970 975
- Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu 980 985 990
- Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn 995 1000 1005
- Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp 1010 1015 1020
- Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg 025 1030 1035 1040
- Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro 1045 1050 1055
- Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
 1060 1065 1070
- Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg 1075 1080 1085
- Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu 1090 1095 1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile 105 1110 1115 1120

- Thr Gly Asp Asn Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
 1125 1130 1135
- Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu 1140 1145 1150
- Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
- Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr 1170 1175 1180
- Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys 185 1190 1195 1200
- Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu 1205 1210 1215
- Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val 1220 1225 1230
- Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu 1235 1240 1245
- Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser 1250 1255 1260
- Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys 265 1270 1275 1280
- Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val 1285 1290 1295
- Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr 1300 1305 1310
- Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln 1315 1320 1325
- Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp 1330 1340
- Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr 345 1350 1355 1360
- Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser 1365 1370 1375
- Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys 1380 1385 1390
- Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val 1395 1400 1405
- Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp 1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu 425 1430 1435 1440

- Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr 1445 1450 1455
- Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 1460 1465 1470
- Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu 1475 1480 1485
- Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys 1490 1495 1500
- Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr 505 1510 1515 1520
- Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
 1525 1530 1535
- Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val 1540 1545 1550
- Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro 1555 1560 1565
- Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro 1570 1575 1580
- Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp 585 1590 1595 1600
- Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg 1605 1610 1615
- Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr 1620 1625 1630
- Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly 1635 1640 1645
- Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg 1650 1655 1660
- Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp 665 1670 1675 1680
- Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly 1685 1690 1695
- Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr 1700 1705 1710
- Lys Leu Lys Leu Thr Pro Ile Ala Ala Gly Gln Leu Asp Leu Ser 1715 1720 1725
- Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val 1730 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Ala 745 1750 1755 1760

- Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro 1765 1770 1775
- Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp 1780 1785 1790
- Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu 1795 1800 1805
- Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser 1810 1815 1820
- Glu Arg Ser Gln Pro Arg Gly Arg Gln Pro Ile Pro Lys Ala Arg 825 1830 1835 1840
- Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu 1845 1850 1855
- Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg 1860 1865 1870
- Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg 1875 1880 1885
- Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu 1890 1895 1900
- Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg 905 1910 1915 1920

Ala